A Novel Method for Measuring the Timing Website Reconstructions

Al Montaser, T.

DOI 10.1016/j.mex.2021.101438
ESSN 2215-0161

Publisher: Elsevier

MethodsX 8 (2021) 101438

Contents lists available at ScienceDirect

MethodsX

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

Method Article

A novel method for measuring the timing website reconstructions

Tariq S. Almontaser Dr, PhD [a,b,1]

[a] Coventry University, Faculty of Busness and Law, The department of Economics, Finance and Accounting, Coventry, UK
[b] The College of Applied Administrative and Financial Sciences Libya, Tripoli, Libya

A B S T R A C T

This research method contributes to the literature by measuring commensurably *the Timing of Website Reconstructions* (TWR) for Web-based research studies. TWR deepens the understanding the functionality of online media by focusing on specific web metrics to make the usage of this type of media efficient for disclosure function. This method helps establishing a new methodological pathway by providing chronological map of website reconstructions. The inputs of new method are originally based on publicly available data and it can be applied through 3 consecutive steps. Accordingly, this method is applicable to assess disclosure function for any web-based research study. In sum, this method presents.

- The inputs of TWR method are publicly available data.
- This method is applicable for variety web-based research studied regardless of the applied methodological approach (e.g., qualitative, quantitative).

A R T I C L E I N F O
Method name: The Timing Website Reconstructions (TWR)
Keywords: Web-based Disclosure, Launching webpages, Web metrics, Online Reporting
Article history: Received 9 April 2021; Accepted 1 July 2021; Available online 2 July 2021

E-mail addresses: tariq@almontaser.com, ad5854@coventry.ac.uk
[1] Associate Lecturer in Accounting - Coventry University - Faculty of Business and Law - Department of Economics, Fiannce and Accounting.

https://doi.org/10.1016/j.mex.2021.101438

Specifications table

Subject area:	Economics and finance
More specific subject area:	*All web-based research studies.*
Method name:	*The Timing Website Reconstructions (TWR)*
Name and reference of original method:	The original method has no name. Reference: Murphy, J., Hashim, N. H., and O'Connor, P. (2007). Take me back: Validating the wayback machine. *Journal of Computer-Mediated Communication, 13*(1), 60–7
Resource availability:	-

Method overview

The extant of web-based research studies are lacking historical data related to the reconstruction of website hierarchies. Thus far, website's hierarchy is generally not a variable of interest in the current literature. Few research studies considered, for example, the importance of presenting a 'site map' on websites—which is based on website construction—as an indicator for effective use of the web for disclosure [2]. Such research approaches focus on observational indicators of the studied websites to assess the visibility of disclosures. While scholars stress the significance of hierarchical structure of webpages [7], the new method helps researchers in assessing the usage of this type of media from different perspectives.

Different features and properties in different medium might lead to variant ways of disclosing information and communication with stakeholders. Consequently, the efficiency of using each type of medium should be evaluated differently according to not only the specifications of used media tools but also the extent of proficiency of using them. Updateability of websites is not just frequent purposeful changes in web content. Updating Web content reflects, for example, the awareness of improving communication with stakeholders [5,6], supporting firms' transparency and enhancing corporate reporting [1].

While this method can be utilized in any disclosure-based research study, the new research method shall be illustrated as an application of CSR area. The sample data has been collected from websites of oil companies operating in Libya between 2008 and 2015. TWR method is based on identifying the popularity dates of birth—which reflects ages—of web pages, where information is published, is methodologically associated with stakeholders' engagement in the contents of which those web pages contain. In other words, the more the popular CSR webpage the more the stakeholders' usage and, probably, engagement. One of the applications of TWR method is translating chronologically the probable changes in firms' agenda through the extent of website reconstructions took place (or launching individual web pages) during a period of time.

The inputs of TWR method

Crawling chronological data about websites' reconstruction is the challenge of many researchers. The two available sources of such information are Google and Archive.org. Table 1 shows that the extent of responsivity of the two sources to launching 30 different websites. Accordingly, Archive.org provides historical information of websites' reconstruction timing more precise than Google. This website provides chronological information about 'captured' websites that have been archived [4].

The archiving mechanism of this website is the more the 'popular' website, the greater the responsivity to its updates[1]. Archive.org provides history-website-version map of millions of websites. Fig. 1 shows an example of history-website-version map with timing details between two dates. I called the first date as 'the popularity date of birth' and the recent data is 'the date of the end of popularity life'. The difference between these two dates is an equivalent to 'popularity life'[2].

[1] See https://archive.org/about/faqs.php#The_Wayback_Machine, accessed in 15/02/2016.

[2] http://en.noclibya.com.ly/index.php?option=com_content&task=view&id=1350&Itemid=55. It should be prefixed with http://web.archive.org/web/*/ .

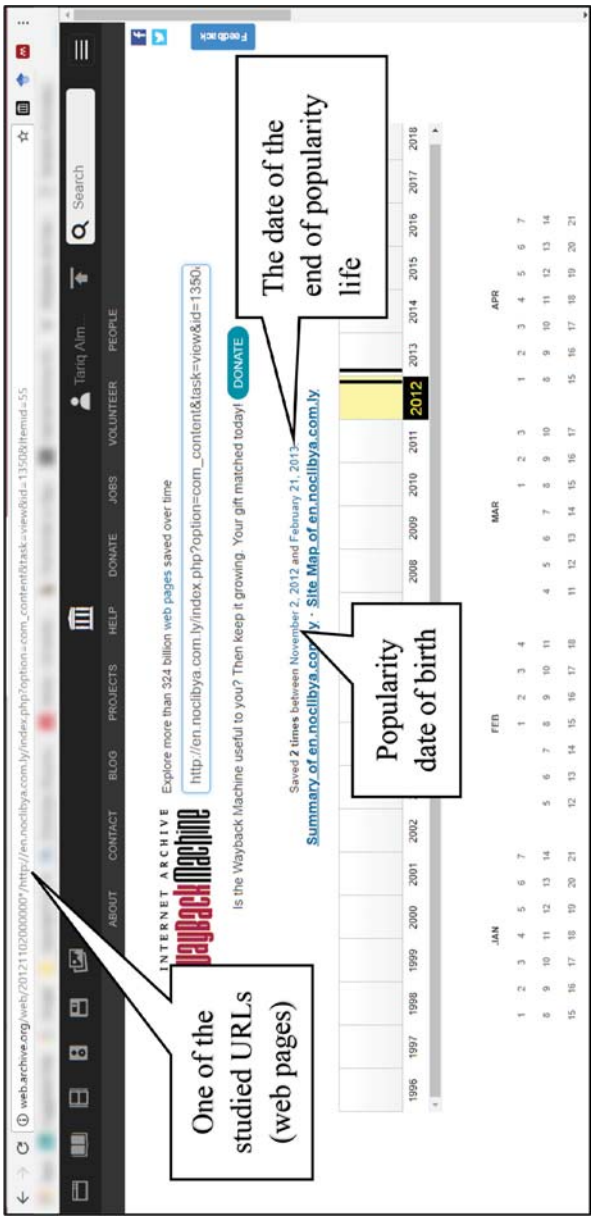

Fig. 1. The illustrations of how timing data of CSRWR can be collected (this is an example of one sample web page).

Table 1
The Estimated dates of launching the websites of a set of global and local (Libyan) oil companies (conducted in March 2018).

URL of the Website	Is itGlobal?	The date of oldest version of the website	
		Archive.org	Google
https://noc.ly	No	26/04/2004	**01/10/2002****
https://www.akakusoil.com	No	**28/06/2008**	25/07/2017
https://sirteoil.com.ly	No	07/02/2011	**27/09/2010**
https://www.zueitina.com.ly	No	**13/12/2007**	27/08/2009
https://www.retajoil.com	No	**07/06/2011**	25/04/2015
https://www.akeel-corp.com	No	02/02/2011	**07/02/2010**
https://adwoc.com	No	**25/03/2003**	Not found*
https://www.wahaoil.net	No	**03/02/2006**	15/06/2009
https://www.vebaoil.com.ly	No	**17/10/2007**	Not found
https://www.vebalibya.com	No	**28/03/2003**	18/03/2008
https://www.harouge.com	No	23/07/2008	**18/03/2008**
https://www.mellitahog.ly	No	01/09/2009	**04/11/2008**
https://www.brega-ly.com	No	**07/11/2004**	Not found
https://alrahila.ly	No	**31/08/2008**	17/09/2008
https://arc.com.ly	No	**10/03/2008**	24/09/2014
http://www.opsint.com	No	**29/12/2001**	Not found
http://www.emacogroup.eu	No	**08/08/2009**	Not found
http://www.crosco.com	No	11/11/1998	**17/12/1996**
http://www.repsol.com	Yes	16/10/1997	**15/04/1997**
http://bosna-s.ba	No	**16/10/2003**	14/06/2006
http://www.africa-oil-gas.com	No	**12/12/2008**	17/01/2009
http://www.eni.it	Yes	**26/12/1996**	04/06/2016
http://www.medservmalta.com	No	**07/03/2004**	10/01/2013
http://www.slb.com	Yes	**06/12/1996**	15/02/1999
http://www.bp.com	Yes	**20/10/1996**	01/04/1997
http://www.total.com	Yes	**06/04/1997**	01/03/2003
http://www.suncor.com	Yes	**26/12/1996**	20/10/2003
http://www.hess.com	Yes	**21/12/1996**	01/02/2001
http://www.marathonoil.com	Yes	**22/12/1996**	05/03/2009
http://www.omv.com	Yes	**07/12/1998**	15/03/2006
The number of oldest dates (out of 30)		23	7
The availability of data		100%	83%
The logarithmic average of dates		**17/10/2003**	04/03/2007

* No search results using Google Search Engine.
** **Bold** dates are the oldest and, consequently, more accurate.

History-website-version map of each URL over all data records of the studied phenomena (e.g., CSR disclosure) must be checked using this prefix-URL[3]. Collecting the 'popularity date of birth' in each data record is achieved through the steps illustrated in Fig. 1.

Adjusting the outputs of TWR method

The outputs of research method should be structured and presented using logarithmic averages of the timing data of *launching* web pages which is different from the timing data of *publishing* CSR data records of which the *launched* web pages are (or *not*) permanently consisted during their popularity life. This means that not all *launched* web pages in specific period are vehicles for all (or some) data records *published* during the same period and vice versa.

I developed this method from Murphy et al. [4] who are the first introduced it to determine ages of websites but without testing its validity and reliability of the used method. The introduced research aims to identify the gap between the 'birth date of popularity' of web pages and the 'actual date of launching' them. The analytical test shall be conducted on the collected data to commensurably *adjust* the 'birth date of popularity'. In other words, this test aims to identify the responsivity of Archive.org

[3] http://web.archive.org/web/*/<URL of data record>

Table 2
The results of analytical test - the average of differences between and Publishing Dates (PDs) of CSR information and the Dates of the First Next Capture (DFNCs) of the same web pages that have been occupied exclusively for CSR information.

Description	2008–2009	2010–2011	2012–2013	2014-2015	All 8 years
A- The number of all data records	529	479	474	519	2001
B- The number of data records have PDs (1st criterion)	383	282	269	373	1307
C- Percentage of data records with PDs (**B** ÷ **A**)	72%	59%	57%	72%	65%
D- The number of data records have PDs (1st criterion) *and* found on web pages used exclusively for CSR information (2nd criterion).	119	256	150	183	708
E- % of data records with 1st and 2nd criteria (**D** ÷ **A**)	22%	53%	32%	35%	35%
Results of Analytical test:					
F- Max differences between DFNCs and PD (in months)	58	76	42	21	76
G- Min differences between DFNC and PD (in months)	0	0	0	0	0
H- The percentage of the number of zero-difference data records out of all data records (when PDs = DFNCs)	1.1%	1.5%	19%	23%	11%
I- Avgs. of the differences between DFNCs and PDs (in months)	13	8	8	6	8

to this context (oil companies working in Libyan oil sector) in terms of capturing the updates of their websites. As mentioned before, the mechanism of Archive.org is based on how popular a web page is (or a website) compared to others as stated on Archive.org "*Internet Archive's crawls tend to find sites that are well linked from other sites*"[4]. The popularity of web page (or website), accordingly, may be attributed to the context and industry to which (large or small) number of stakeholders are attractive; the population of internet society in the context could perhaps be an influential factor [3].

To identify the responsivity of Archive.org to, for example, Libyan context, the statistical comparison shall be made in each data record (web page) between the '*Dates of Publishing CSR information*' (DPs) and the *Dates of the First Next Capture* (DFNCs) of the same webpage. This could be an appropriate test to determine how the actual dates of launching web pages, which is unknown, in each period are averagely earlier than 'the birth dates of popularity' of web pages during the same period.

However, not all oil companies date their posts from where we captured CSR information. As shown in Table 2, item C, two third (63%) of data records of studied data that have been published with dates of posting CSR information on website is the first criterion. Regardless, CSR data records related to the disclosure dated by oil companies are only considered in analytical test of this method.

To minimise the probable bias in the results of analytical test, the web pages have been occupied exclusively by CSR & sustainability information are only considered in this test as the second additional criterion (see Table 2, item E). Accordingly, the analytical test is conducted by the third of data records.

In general, the results of analytical test of CSRWR method shows that *the dates of the next capture* are averagely 8 months after *the dates of publishing* as presented in Table 2. This implies that *the actual dates of launching* web pages are 8 months earlier than *the birth dates of popularity.* However, the adjustment of each period should be used to the adjust *the birth dates of popularity* of each data record by taking off 'the average of the differences between DFNCs and PDs' of that period, which is shown on Table 2. For example, if *the birth date of popularity* of URL of specific data record is in December 2014, the estimated date of launching this webpage is December 2014 *minus* 6 months = June 2014.

The analytical test reveals either the rapid increase of significance of Libyan context in capturing progress of Archive.org or the rapid increase of the general capability of Archive.org per se over time. The lag of responsivity has declined from 13 months in 2008-2009 to just 6 months in 2014-2015. However, these results are incomparable with responsivity of Archive.org to the updates of, for

[4] https://archive.org/about/faqs.php, accessed in 01/07/2016.

Table 3
The average dates (quarterly basis) of launching new webpages of oil companies regarding different types of ownerships and regardless of website versions.

Types of Ownership	Pre-revolution era		Post-revolution era		Average	# of launched webpages
	2008–2009	2010–2011	2012–2013	2014–2015		
Owner	⬅ Q2-2007	Q1-2010	⬅ Q2-2011	Q3-2014	**Q4-2010**	543
JVCs	⬅ Q3-2007	⬅ Q1-2009	⬅ Q1-2011	Q1-2014	**Q3-2010**	142
NCs	⬅ Q2-2007	⬅ Q4-2008	Q1-2012	⬅ Q4-2013	**Q1-2012**	190
Average	⬅ **Q2-2007**	⬅ **Q4-2009**	⬅ **Q3-2011**	**Q2-2014**	**Q1-2011**	
# of launched webpages	113	382	182	198		**875**

JVCs = Joint Venture Companies; NCs = National Companie

example, BBC website[5] which has been captured once each 3 to 4 hours in 2014! so, there would not be a gap between BDPs and actual dates of launching web pages in such cases.

The application of TWR method

As mentioned before, the introduced research method has been utilised in this paper by CSR data collected from websites of oil companies from 2008 to 2015. The chronological delay/gap of CSRWR has been highlighted in Table 3, by thick left-arrow ' ⬅' as a prefix of the figures concerned. This arrow indicates to the previous period when most used web pages during the current period launched. For example, 2008-2009 witnessed a CSRWR-delay of the owner as presented in Table 3. Most of the CSR information in that period have been published on web pages launched during the second quarter (Q2) of 2007. Accordingly, ' ⬅' highlights the CSRWR-delay of this firm during 2008-2009.

Conclusion

This paper introduced the steps to be followed use TWR method for ant web-based research studies. TWR method helps to commensurably determine the changes in website hierarchies of the studied firms which benefits anyone interested in conducting a research regardless the adopted methodology.

The limitations of TWR method

The outputs of analytical test for adjusting *the birth dates of popularity*, on which research method is based, has been deductively determined. Adopting alternative history-website-version source of data might mitigate the error of estimation. Unfortunately, real timing data of launching webpages/websites is not available by the date publishing this paper. Accordingly, the outputs of analytical test of research method cannot be precisely judged. In other words, the estimation error of that test has not been measured yet.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[5] https://web.archive.org/web/20140801000000*/www.bbc.com , Accessed in 14-05-2018.

References

[1] H. Ashbaugh, K.M. Johnstone, T.D. Warfield, Corporate reporting on the internet, Account. Horizons 13 (3) (1999) 241–257.
[2] G.M. García, M.V. Carrillo-Durán, J.L. Tato-Jimenez, Online corporate communications: website usability and content, J. Commun. Manage. 21 (2) (2017) 140–154.
[3] Information Geographies at the Oxford Internet InstituteInternet Population and Penetration, 2013 Retrieved from http://geography.oii.ox.ac.uk/wp-content/uploads/2013/06/InternetPopulation2011_HexCartogram_v7-01.png .
[4] J. Murphy, N.H. Hashim, P O'Connor, Take me back: validating the wayback machine, J. Comput. Mediated Commun. 13 (1) (2007) 60–75.
[5] N. Waweru, M. Mangena, G. Riro, Corporate governance and corporate internet reporting in sub-Saharan Africa: the case of Kenya and Tanzania, Corporate Governan. 19 (4) (2019) 751–773.
[6] D. Wheeler, J. Elkington, The end of the corporate environmental report? Or the advent of cybernetic sustainability reporting and communication, Bus. Strategy Environ. 10 (1) (2001) 1–14.
[7] E. Wong, R. Leung, R. Law, How effective are asian hotels in communicating csr efforts through the property websites? the case of Hong Kong, in: I. Tussyadiah, A. Inversini (Eds.), Information and Communication Technologies in Tourism 2015, Springer International Publishing, 2015, pp. 651–663.