

Monitoring Online User Behaviour. The case of the Newstracker

Martijn Kleppe

Vrije Universiteit

m.kleppe@vu.nl

Irene Costera Meijer

Vrije Universiteit

lcostera.meijer@vu.nl

Digitalization of journalism enables news organizations to monitor the behavior of online news users by using metric tools such as Google Analytics. Journalists and news executives take these data as good measurements of audiences' interest in news, but they cover only their own websites. To create a full picture of all online news consumption, research companies create lists of aggregated visits to websites. However, this type of research does not consider website visitors as citizens but as commodity (Richardson, 2007, p. 79; Usher, 2013), giving advertisers detailed information on how to reach their target audience in the most efficient manner. Feedback to journalists is limited to presenting the most clicked items, often leading to the critique that most news users are more interested in trivial news than in public affairs (Boczkowski, Mitchelstein, & Walter, 2011; Karlsson & Clerwall, 2013; Nederlandse Nieuwsmonitor, 2013). Relatively little is known about the content of the visited news articles, let alone the everyday patterns of individual news consumption (Costera Meijer & Groot Kormelink, 2014).

Therefore, this paper will present the set-up and first results of the research tool 'Newstracker' developed especially for monitoring news consumption from the angle of the user. We first installed a proxy on the laptop of a group of 50 respondents who use this device regularly for the consumption of news and information. This set-up is in line with Findahl (2013) who investigated the online behaviour of an American family and Menchen-Trevino (2012) that used a special-designed proxy to monitor the exposure to political communication during the November 2010 U.S. general election campaign. However, these academic studies only report the website titles that have been visited. Our set-up goes two steps further. We do not only monitor the website titles but also the actual visited URLs and we crawl all textual and visual contents of the visited websites. Since one of the problems when monitoring a person's online behaviour is the magnitude of the data that is being collected (Batista & Silva, 2002; Manovich, 2012; Vicente-Marino, 2013, p. 43), we deploy automated content analyses techniques (Atteveldt, 2008; Bhulai, Kampstra, Kooiman, Koole, & Kok, 2012) to detect the topics that are being discussed in the news items. This enables us to calculate the topical online news consumption during the day.

In this paper we will discuss the set-up of our research tool and its applicability for other types of Digital Humanities research such as user studies focussing on formulating requirements , based on existing user behaviour.

References

- Atteveldt, W. van, 2008. *Semantic Network Analysis: Techniques for Extracting, Representing, and Querying Media Content*. BookSurge.
- Batista, P., Silva, M., 2002. Mining Web Access Logs of an On-line Newspaper. Presented at the Proceedings of the Workshop on Recommendation and Personalization in eCommerce of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based Systems, Malaga.
- Bhulai, S., Kampstra, P., Kooiman, L., Koole, G., Kok, B., 2012. Trend visualization on Twitter: What's hot and what's not In: *Data Analytics*. Presented at the IARIA, Barcelona, pp. 43–48.
- Boczkowski, P.J., Mitchelstein, E., Walter, M., 2011. Convergence Across Divergence: Understanding the Gap in the Online News Choices of Journalists and Consumers in Western Europe and Latin America. *Commun. Res.* 38, 376–396. doi:10.1177/0093650210384989
- Costera Meijer, I., Groot Kormelink, T., 2014. Checking, Sharing, Clicking and Linking. *Digit. Journal.* 0, 1–16. doi:10.1080/21670811.2014.937149
- Findahl, O., Lagerstedt, C., Aurelius, A., 2013. Triangulation as a way to validate and deepen the knowledge about user behavior. A comparison between questionnaires, diaries and traffic measurement, in: *Audience Research Methodologies: Between Innovation and Consolidation*. New York, pp. 54–69.
- Karlsson, M., Clerwall, C., 2013. Negotiating Professional News Judgment and “Clicks.” *Nord. Rev.* 34, 65–76. doi:10.2478/nor-2013-0054
- Kleinneijenhuis, J., Atteveldt, W. van, 2006. Geautomatiseerde inhoudsanalyse, met de berichtgeving over het EU-referendum als voorbeeld, in: *Inhoudsanalyse: Theorie En Praktijk*. Kluwer, pp. 227–250.
- Manovich, L., 2012. How to Follow Software Users?
- Menchen-Trevino, E., Karr, C., 2012. Researching Real-World Web Use with Roxy: Collecting Observational Web Data with Informed Consent. *J. Inf. Technol. Polit.* 9, 254–268. doi:10.1080/19331681.2012.664966
- Nederlandse Nieuwsmonitor, 2013. Seksmoord op horrorvakantie: de invloed van bezoekersgedrag op krantenwebsites op de nieuwsselectie van dagbladen en hun websites.
- Richardson, J.E., 2007. *Analysing newspapers: an approach from critical discourse analysis*. Palgrave Macmillan, Basingstoke [England]; New York.
- Usher, N., 2013. Al Jazeera English Online. Understanding Web metrics and news production when a quantified audience is not a commodified audience. *Digit. Journal.* 1, 335–351. doi:10.1080/21670811.2013.801690
- Vicente-Marino, M., 2013. Audience research methods. Facing the challenges of transforming audiences, in: *Audience Research Methodologies: Between Innovation and Consolidation*. New York, pp. 37–53.