What do we know about Altmetric.com sources? A study of the top 200 blogs and news sites mentioning scholarly outputs

Fraumann, Grischa; Zahedi, Zohreh; Costas, Rodrigo

Veröffentlichungsversion / Published Version
Sonstiges / other

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-72013-9

What do we know about Altmetric.com sources?

A study of the top 200 blogs and news sites mentioning scholarly outputs

Grischa Fraumann[1, 2], Zohreh Zahedi[1], Rodrigo Costas[1]

grischa_fraumann@gmx.de; z.zahedi.2@cwts.leidenuniv.nl; rcostas@cwts.leidenuniv.nl

1 CWTS, Leiden University, P.O. Box 905, Leiden, 2300 AX (The Netherlands)

2 University of Tampere, School of Management / Higher Education Group, FI-33014, Pinni A4022, Kanslerinrinne 1, Tampere (Finland)

Keywords

Introduction, Data and Methodology

This paper presents a preliminary study of the diversity and typology of users and uses around scholarly outputs in blogs and news sites as tracked by Altmetric.com. The top 100 blogs and top 100 news websites in terms of mentioning publications, for which their URLs are available, have been considered for a deeper analysis. The identified sources were manually analyzed on the respective websites and on the available Altmetric.com metadata and classified in order to understand how the scholarly outputs are mentioned (table 1).

Categories	Sub categories and percentage of all retrieved sources (n=200)	Top 100 Blogs[1]	Top 100 News
Activity	• Active / not active (not active refers to sites were the last entry occurred in 2014) active = 79%; inactive = 6.5% • Broken link (content not available) = 3 % • Duplicate (similar or same link/source in the database) = 11.5%	79 active 13 inactive 5 broken links 3 duplicates	79 active 20 duplicates 1 broken link
Language (Main) languages of the websites	• English = 80.5% • German = 3% • Italian = 0.5% • Japanese = 1% • Norwegian = 0.5% • Polish = 0.5% • Portuguese = 0.5% • Russian = 0.5% • Spanish = 1.5% • Turkish = 0.5%	88 English 2 Spanish 2 Japanese 2 German 1 Italian 1 Turkish 1 Russian 1 Portuguese 1 Polish	73 English 4 German 1 Norwegian 1 Spanish
Engagement	• Introduction (short explanation of the topic) = 12.5% • Discussion (consideration and examination of an issue) = 71.5% • Link (to articles) = 82.5% • Title (of the article) = 74.5% • Source (of the article) = 79% • Complete bibliographic information of the article = 26% • Unknown = 1%	89 links 85 titles 82 sources 67 discussions 38 complete bibliographic information 22 introductions 1 unknown	76 discussions 76 links 76 sources 64 titles 14 complete bibliographic information 3 introductions 1 unknown
Country of the websites	• Australia = 2% • Brazil = 0.5% • Canada = 1.5% • Germany = 2% • Hong Kong = 0.5% • India = 0.5% • Italy = 0.5% • Japan = 1.5% • Monaco = 0.5% • Norway = 0.5%	US: 61 UK: 4 Japan: 3 Australia: 2 Canada: 2 Italy: 1 Monaco: 1	US: 65 Germany: 4 UK: 4 Australia: 2 Canada: 1 Spain: 1 India: 1 Norway: 1 South Africa: 1 Hong Kong: 1 Brazil: 1

---

[1] The overall sample of blogs and news sites (n=200) is not always reached in this table, because of multilingual sources, several countries, broken links, and duplicates among the retrieved URLs.

Categories	Sub categories and percentage of all retrieved sources (n=200)		Top 100 Blogs[1]	Top 100 News
	• South Africa = 0.5% • Spain = 0.5% • UK = 4% • US = 63%			
Authors	**Blogs**	**News**	52 institutional 36 individual authors 19 professors 11 researchers 5 unknown 1 layperson/citizen scientist	Journalists[2]
	according to the information on the blog • institutional (usually several authors affiliated to a media institution that provides the blog) • individual author(s) (one or a few authors *not* affiliated to a media institution that provides the blog) • professor(s) (at a university, research institute, or similar) • researcher(s) (not closer specified academic affiliated to an academic organization) • layperson(s)/citizen scientist(s) • unknown	journalists[3]		

Table 1: Top 100 blogs and Top 100 news sites: summary of classification and results

---

[2] No further sub categories were introduced in this case. During the analysis, the categorization of the news site as such became more important than the one of the journalists. This is based on the fact, that the news articles by several journalists of one news site follow mostly similar standards.
[3] There might be some overlap of this subcategory with blogs, as some blog authors can also be seen as journalists.

Table 1 shows that mainly English-speaking sources based in the United States (US) are included in the global Altmetric.com top 200 sources (61 blogs and 65 news sites). In addition, two odd[4] blogs were spotted among all 200 sources.

Preliminary results and discussion

In the dataset there were only 5 broken links of blogs and 1 broken link of news sites. In total, 20 duplicate news sites, and 3 duplicate blogs were retrieved and 3 sites were listed in Altmetric.com database as blog and as news site at the same time. Regarding engagement, there are differences between news and blogs. Some blogs provide only introductions (22), whereas this rarely happens on news sites (3), which provide more context. 38 blogs provide complete bibliographic information about the publications they are discussing.

Most blogs provide the links to the publications (89), then the titles (85) and the sources (82). They present a discussion (67) and as mentioned above 22 provide an introduction. News sites provide substantial discussions (76) and less bibliographic information (14). 76 news sites provide links and sources, 64 mention the titles of the publications.

Blogs and news sites in the dataset share similar characteristics such as author names, providers and thematic focus. The most visible difference is the difference in providing a complete bibliographic information when the blog or news author refers to scholarly output. Some sites are part of the same media institution, (e.g. a publisher) for specialized blogs or localized news with sub domains. Some blogs are not active anymore, whereas almost all news sites are active. 52 blog authors are affiliated to media institutions; the record shows further 36 individual authors, 19 professors, 11 researchers, and 1 layperson/citizen scientist.

Some news items are posted on several news sites, thus duplicating the counted mentions. This might be a common practice in news media, especially regarding press releases, but it is important to further understand the effects that this may have in calculating indicators.

---

[4] One is http://sciencealerts.com/ which does not allow access to the blogposts anymore and the title of the blogposts would fully coincide with that of the article, and http://organometallics.blogspot.ca/ which presents images and links to chemistry publications but without further discussion.

Conclusion

This paper presents a preliminary analysis of two important altmetric sources: blogs and news media mentions. Issues identified in this study that would need further discussion include the presence of duplicates across several sources; the identification of suspicious sites, etc. By properly understanding the types of sources and users behind the collected altmetric data, one might more easily relate the actual impact and value of mentioning the publications.

Further research will consider the enlargement of the studied sources and other source types. Other altmetric providers could also be included. The analysis could be expanded with a text mining of the different sites, to extract frequent keywords of the posts or even the comments by site visitors. Future studies could concentrate specifically on online social networks (OSN).