Mining "big data" using big data services

Reips, U.D.; Matzat, U.

Published in:
International Journal of Internet Science

Published: 01/01/2014

Document Version
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 18. Dec. 2018
Mining “Big Data” using Big Data Services

Ulf-Dietrich Reips¹, Uwe Matzat²

¹University of Konstanz, Germany;
²Eindhoven University of Technology, The Netherlands

Big Data on bare feet: Free tools for analysis

While many colleagues within science are fed up with the “big data fad”, empirical analyses we conducted for the current editorial actually show an inconsistent picture: we use big data services to determine whether there really is an increase in writing about big data or even widespread use of the term. Google Correlate (http://www.google.com/trends/correlate/), the first free tool we are presenting here, doesn’t list the term, showing that number of searches for it are below an absolute minimum that is even mastered by terms like “brobdingnagian” or “sockdolager” (which apparently correlates >.80 in search patterns with searches for “oriental princess” btw.).

Figure 1. Searching for “big data” in tweets with iScience Maps Global search (http://tweetminer.eu).

Address correspondence to Ulf-Dietrich Reips, Psychological Methods, Assessment, and iScience, University of Konstanz, Box 31, 78457 Konstanz, Germany, reips@uni-konstanz.de. We acknowledge support by COST Action 1004 “WEBDATANET” (http://webdatanet.eu).
On the other hand, people do tweet about big data, and we use our tool iScience Maps (http://tweetminer.eu, Reips & Garaizar, 2011) to do a global search for “big data”. Figure 1 shows a world map produced by the tool, showing the areas where most tweets originated from a random sample of about 5-10% of all tweets between 2010 and now. Each dot on the map can be clicked, showing further detail of the area, in the example case Palm Beach County in Florida. From there we can inspect each tweet for that geo-location on Twitter, see Figure 2, and the many results convince us there are big data in tweets.

Using Google Books NGram Viewer (http://books.google.com/ngrams) we can compare frequencies of appearance of terms in books. Choosing the US English digitized corpus of books–big data!–we search for appearances of our target term during the years between 1980 and 2008 (newest year possible in NGram Viewer). Indeed, albeit having peaked in 2001, “big data” seems to fare well compared to “pig data”, see Figure 3. But doubts remain. Welcome to the wonderful new world of pig data mining!

![Figure 2. Tweets containing the term “big data”, identified by geolocation.](image)

![Figure 3. Identifying the frequency and timeline of appearance of the terms “big data” and “pig data” in books via Google Ngram.](image)
Updated figures for the International Journal of Internet Science

The International Journal of Internet Science continues to be under evaluation for inclusion with the ISI Web of Science database. You may have seen from our editorials in issues 3(1), 6(1), 7(1), and 8(1) that calculations result in high journal impact. We will show below that the journal continues to be an increasingly important source for researchers.

Figure 4. Web page for the International Journal of Internet Science in Google Scholar, available at http://scholar.google.com/citations?hl=en&user=OCYy1o4AAAAJ

Our journal page on Google Scholar Citations (see Figure 4) at http://scholar.google.com/citations?hl=en&user=OCYy1o4AAAAJ allows one to track citations to the journal and its articles, link as an author, and enter one’s e-mail address for automatic notifications. When we calculated the impact in late December 2011 Google Scholar showed 198 citations (Reips, 2011), in June 2013 there were 441 (Reips & Matzat, 2013), now there are 703. The three most frequently cited articles are by Göritz (2006), Freelon (2010), and Mesch and Talmud.
(2006), with 168, 104, and 51 citations, respectively. The most popular recent item, our 2012 editorial on Big Data (Snijders, Matzat & Reips, 2012), has already received 31 citations.

**IJIS’ 2013 impact: > 6.1 (ISIIJIS not included > 1.3; ISIIJIS included > 2.0)**

The ISI journal impact is calculated as the average number of citations in a year given to those articles in a journal that were published during the two preceding years (Impact factor, 2011). The most current journal impact that can be calculated is thus the one from 2013. We therefore need to count all citations in 2013 to citable items (articles, editorials, book reviews etc.) that appeared in the International Journal of Internet Science during the years 2012 and 2011. These citations are listed in the following paragraph.


Thus, the currently official 2013 journal impact calculation is $43 / 7 = 6.1$. If we calculate only with citations that appeared in ISI journals (Thomson only includes journals in its calculations that are already in its database), then we arrive at an **ISI impact of 9 / 7 = 1.3** for the International Journal of Internet Science. If we calculate as if the International Journal of Internet Science was already part of the ISI database, then its **ISI impact is 14 / 7 = 2.0**, because our 2013 editorial cites five citable items from the International Journal of Internet Science. Importantly, all figures calculated here are conservative estimates, because more citing articles published in 2013 are not yet known and couldn’t be included in the present analysis. While the 2013 figure has not changed from the 2012 figure for Google Scholar, but is down from 2012 for the ISI figures (Google Scholar: > 6.1; ISIIJIS not included > 2.3; ISIIJIS included > 2.9), due to the remarkably successful Freelon (2010) article only contributing to the latter, we are happy to report that the International Journal of Internet Science’s impact is still on a comparatively very high level and will likely substantially increase again next year because of high citation rates to citable items in 2012 and 2013. Already, there are 43 citations in 2014 (by July!) to citable items in 2012 and 2013. **If they double by the end of the year the 2014 journal impact for the International Journal of Internet Science will be >10.**

The IJIS compares well with other long-standing journals in the field, examples are **Behavior Research Methods** (ISI impact of 1.9), **Computers in Human Behavior** (2.1), **Cyberpsychology Behavior and Social Networking** (1.8), **International Journal of Human-Computer Studies** (1.4), **New Media & Society** (1.8), **Social Science Computer Review** (1.3) and **Information Society** (1.1, all values from the 2012 JCR Social Science Edition).

**The present issue**

In the first article, **Use of the Internet in Capital Enhancing Ways–Ethnic Differences in Israel and the Role of Language Proficiency**, Sabina Lissitsa (Ariel University, Israel) and Svetlana Chachashvili-Bolotin (Ruppin Academic Center, Israel) utilize large scale CBS data to analyse the divide between the digital divide and digital inequalities between immigrants and Jewish veterans in the Israeli society. With respect to differences in access to the Internet they find that important gaps in access disappear after controlling for Hebrew language proficiency. With respect to gaps in engagements in Internet activities that have the potential to increase the user’s social capital, they demonstrate that immigrants, once they have access to the Internet, do not show less engagement. Furthermore, within the group of Internet users, only Western immigrants, but not other groups of immigrants show less engagement in human
capital-enhancing Internet activities when compared to Jewish veterans. The authors discuss the policy implications of their findings.

In the second article, Katharina Hanel and Martin Schulze (both Heinrich-Heine-University of Düsseldorf, Germany), in their article Analyzing the Political Communication Patterns of Voting Advice Application users, examine the use of a popular Voting Advice Application (VAA) by German Internet users. VAAs have recently become popular in Europe. They enable individual Internet users to compare their position on policy issues with those of political parties and candidates running for election. The authors study the communication habits of VAA users and distinguish five classes of political communication patterns in a latent class analysis. Their results show that the probability of VAA use is affected by the breadth of users’ political communication. The German VAA reaches a diverse group of users with regard to political communication, socio-demographic characteristics, and political interests.

Michael Schreiner (University of Education, Heidelberg, Germany), Siegbert Reiss and Karl Schweizer (both Goethe University Frankfurt, Frankfurt a. M., Germany) report on Method Effects on Assessing Equivalence of Online and Offline Administration of a Cognitive Measure: The Exchange Test. In assessing online-offline equivalence articles in the literature often report analyses based on one measurement only or aggregate data from repeated measurements. The present study shows by counterbalanced test-retest-design that there is the possibility of method effects that can only be detected in a cross-mode test-retest design and with appropriate analyses. In a counterbalanced test-retest-design the web-enabled version of the Exchange Test, a cognitive measure of working memory capacity that requires the reordering of four symbols in single exchanges until the sequence of these symbols corresponds to a pre-set sequence of the same symbols, was administered offline and online. Results indicated that both administration forms were parallel at the first measurement. Hence, online and offline administration seemed equivalent. However, data from the second measurement show they were not: Analyses revealed a main effect of repeated measurement and a systematic interaction of “repeated measurement” and “order of administration” (online-offline vs. offline-online). The authors argue convincingly that online-offline equivalence may often only seemingly exist at a superficial level.

Finally, in a non-peer-reviewed supplement, members of the Webdatanet COST network with Innovation and quality in web-based data collection present an update on past and upcoming activities within the network. We encourage you to consider joining the network’s activities, there will be several meetings, training schools, and conferences in locations all over Europe.

Acknowledgements and Contact

The current issue of the International Journal of Internet Science is a result of the devotion, time, and effort of many individuals and institutions that support and help us. Our editorial assistant, Ruoyun Lin (Eindhoven University of Technology, now University of Tübingen), has handled the office work very reliably and professionally. We also thank our outgoing editorial assistant, Dr. Frederik Funke, for providing advice when it was needed.

Grateful acknowledgement goes to the University of Konstanz, the Eindhoven University of Technology, and Webdatanet (http://webdatanet.eu) for their institutional support of and partnership with the journal. And of course, we very much would like to thank the members of IIJS’s Editorial Board and Panel and its many ad hoc reviewers for their valuable contributions to the quality of this journal.

Our editorials in issues 3(1), 6(1), 7(1), 8(1), and the calculations above show very high journal impact that continues to grow. The rejection rate for articles submitted to the International Journal of Internet Science is at 84.1%, meaning that roughly one out of six manuscripts passes the stages of desktop rejection and repeated rounds of multiple peer review.

We encourage our readers to stay in contact with us. Our journal page on Google Scholar Citations at http://scholar.google.com/citations?hl=en&user=OCYy1o4AAAAJ allows you to track citations to the journal and its articles and enter your e-mail address for automatic notifications. Authors can even link themselves in this role. Facebook users may want to subscribe to and “like” our Facebook page at https://www.facebook.com/pages/International-Journal-of-Internet-Science/251579034934885. To get notified of new developments at the International Journal of Internet Science, please subscribe at http://www.ijis.net/subscribe.html.
References


---

1 ISI-listed articles citing articles in the IJIS are marked with an asterisk.


