

Issues in "blogmetrics" - case studies using BlogPulse to observe trends in weblogs

Alastair Smith

AlastairSmith@vuw.ac.nz

School of Information Management, Victoria University of Wellington, PO Box 600, Wellington, New Zealand

Abstract

Weblogs ("blogs") have emerged as a significant communication medium. The BlogPulse Trend Search tool was used for a number of case studies, examining trends in the use of terms in blogs. The paper discusses methodology issues in searching blogs, concluding that blogs can indicate new trends in popular culture, language, and science. However interpretation of trends requires careful examination of the postings, to ensure that an apparent trend is not due to problems in terminology, spam, trivial references to the concepts, etc. In particular, there seems to be no easy way to link postings to particular countries.

Keywords

Weblogs; Search Engines; Longitudinal Studies; Methodology; Case Studies.

Introduction

Weblogs ("blogs") have emerged as a distinct form of communication where non-technical individuals can easily comment on events and issues, and provide links to other blogs. Search engines have emerged that enable searches of blogs to be carried out, raising the possibility of bibliometric research into blogs - "blogmetrics".

This study uses the BlogPulse Trend Search tool (<http://www.blogpulse.com>) to show trends in the use of terms in blogs over time. This paper uses a case study approach to examine the methodology issues that arise in using commercial search engines to investigate blogs. Some questions that are addressed:

- To what extent do terms used in searches reflect the content of the blogs found by the search?
- To what extent are research oriented topics reflected in blog traffic?
- Can national differences be tracked using blog searches?

Literature review

The growth of blogs has lead to a number of research papers on the topic. Reynes-Goldie (Reynes-Goldie, 2004) examined the blog server LiveJournal as a new way of managing information and creating knowledge. Dickie (Dickey, 2004) investigated the use of blogs in distance education to overcome the feelings of isolation experienced by student teachers. The use of blogs by politicians has been studied by Auty (Auty, 2005) and Jackson (Jackson, 2006). Clyde (Clyde, 2004) examined 55 blogs maintained by libraries, and discussed the use of blogs by libraries to communicate with their clients. Bar-Ilan (Bar-Ilan, 2005) monitored a set of blogs over two months, and demonstrated their use as "information hubs" summarising and linking to more substantial information. Herring et al (Herring, Scheidt, Wright, & Bonus, 2005) examined a random sample of blogs, and argued that blogs tended to be individual and intimate, rather than externally focussed. Adar and Adamic (Adar & Adamic, 2005) examined the propagation of information through the blogosphere, using an "epidemic" model. Cohen and Krishnamurthy (Cohen & Krishnamurthy, 2006) carried out a crawl of blogs, identifying the overall size of the blogosphere and its link structure. Delwiche (Delwiche, 2005) compared news stories covered by blogs with those covered by the conventional news media, finding that blogs had moved from technology to broader political issues.

The research indicates that blogs are a growing method of communication, and are worthy of bibliometric investigation.

Methodology

The BlogPulse search engine is described by Glance, Hurst and Tomokiyo (Glance, Hurst, & Tomokiyo, 2004). In the current study, searches on a number of terms that might be expected to show trends in blog discussion were carried out on the BlogPulse Trend Search tool. These produced a graph of the frequency of hits over time. Up to three searches at a time can be compared. It is possible in the BlogPulse Trend Search tool to examine the specific blog postings that gave rise to the hits on a particular date. For this study, a sample of hits were examined to see if the postings were truly reflective of the content expected. Searches were carried out in October/November 2006. Graphs are presented courtesy of Nielsen Buzzmetrics, <http://www.blogpulse.com> who allow use for research purposes.

Results

This section will examine and interpret a number of case study searches. Some representative graphs are included.

Case study 1: blogging a trend

A simple example of the BlogPulse Trend Search tool is to see when a new trend is noticed in the Blogosphere. For example in mid-October 2006 the use of an actress as a search intermediary, Ms Dewey (<http://www.msdewey.com/>), elicited discussion on many blogs, with the peak shown in Figure 1.

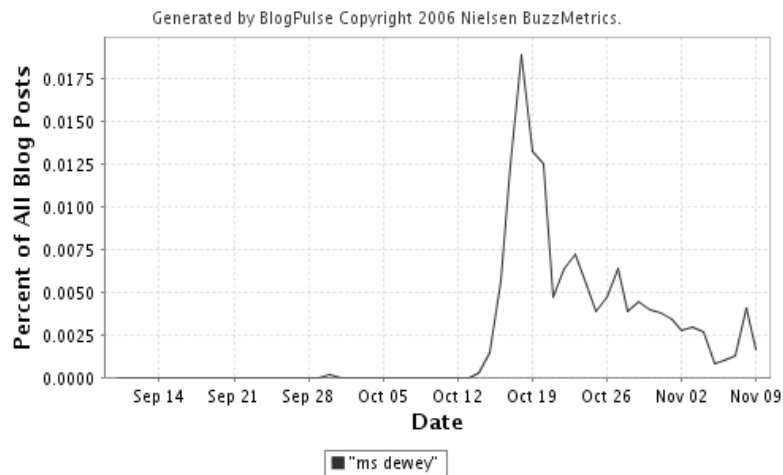


Figure 1: Ms Dewey.

Case study 2: sports terminology

"Football" means gridiron in North America, rugby in Australasia, and soccer in the rest of the world.

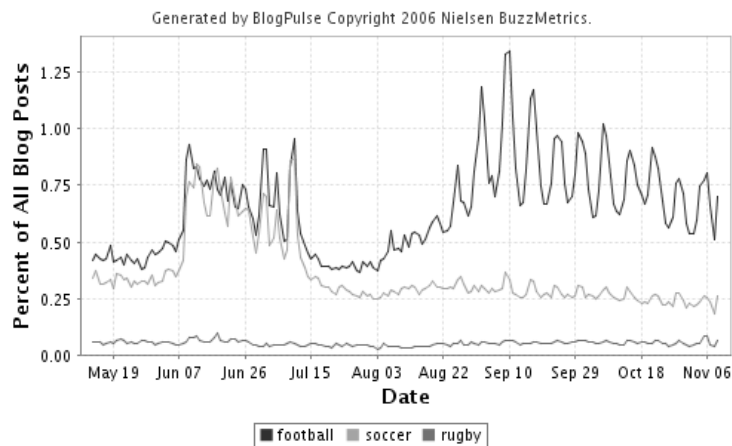


Figure 2: Football

This is illustrated in Figure 2. The June/July peak corresponds to the 2006 FIFA World, so the terms "football" and "soccer" are both used frequently. After this, "soccer" falls off, since it is a minority interest in the countries where the term is used. "Football" continues strongly, although European bloggers are discussing soccer, the North Americans gridiron. In fact many of the references are to football as a background to daily life: "my family sits at home watching football", so the occurrence of a term is not necessarily an indication that it is the subject of the blogging.

Case study 3: Searching for a country's postings

Bloggers tend to use blogging sites which are generally in the .com domain, so the URL gives no indication of the blogger's country. To test if a search on a country name identifies blogs from that country, a phrase search was carried out on three country names: New Zealand, Britain, and United States. A random sample of the first 10 entries were checked.

Table 1: relationship of blog posting to country

Country	Blog originating in country	Tourist account of country	Blogger is expatriate of country	Incidental mention of country	Country is topic
Britain	2	-	-	7	1
New Zealand	2	1	1	7	1
United States	9	-	-	-	1

For Britain and New Zealand, most postings were to incidental mentions for example "Things I'd like to do: Travel: where? Japan, Scotland, New Zealand". The differing United States result reflects the fact that most bloggers are based in the US. The use of a country name in a blog posting is not an indication that the posting is related to that country.

Case study 4: universities' profile in the blogosphere

The names of three NZ universities were searched in Blogpulse (Figure 3). VUW has more mentions than Auckland or Canterbury Universities, although Auckland and Canterbury are larger universities. However this is misleading: the 4 September spike has many postings related to the death of Steve Irwin, born in Victoria, Australia. Also, a number of "victoria university" postings relate to other institutions, for example in Victoria Canada.

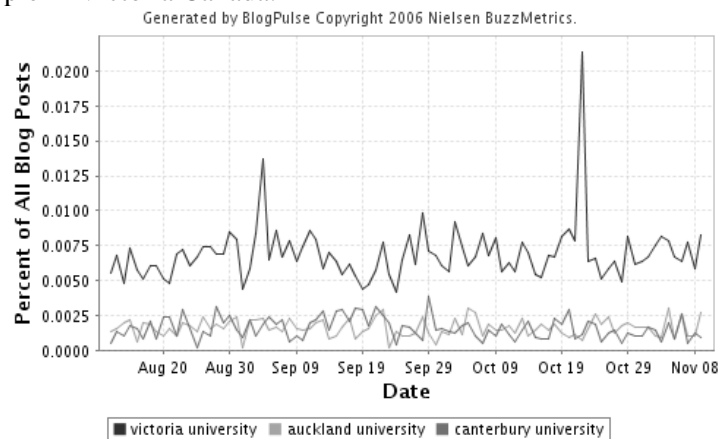


Figure 3: NZ Universities

Case study5 : triple helix in the blogosphere?

The triple helix of University, Government, and Industry has been discussed in the bibliometric literature (e.g. Leydesdorff & Curran, 2000). A search (Figure 4) appears to show a spike at 15 October in discussion of university/government and university/industry relations. However they are spam - spurious blogs set up with random phrases that on this day happen to include "government" and

"university". A few days later the spike had disappeared both from Blogger (the host) and from BlogPulse.

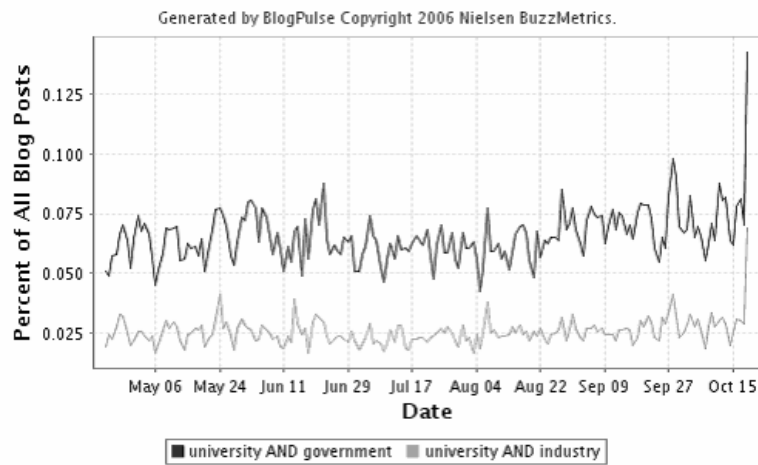


Figure 4: Triple helix

Case study 6: science in the blogosphere - stem cell research

Figure 5 shows very clear peaks related to news of embryonic stem cell research (banning of US federal funding, an actor's appeal for treatment, controversy in the US elections).

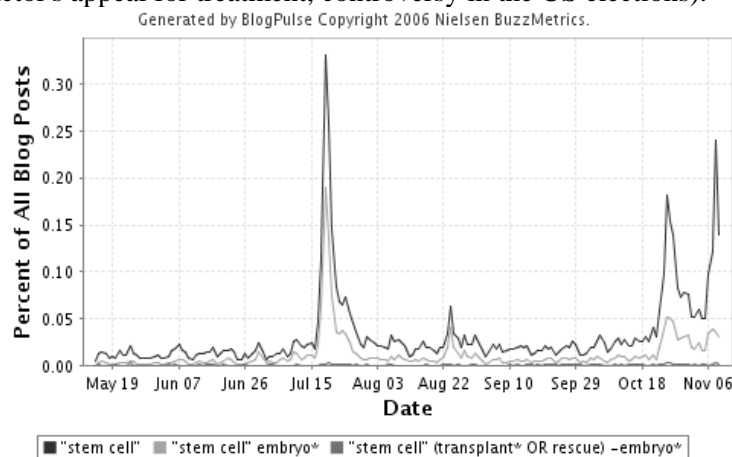


Figure 5: Stem Cell Research

However there is a non-controversial use of stem cells, from donors or from patients themselves, in transplant/rescue treatments of cancer. While the peaks relate to controversial embryonic stem cell use, a steady but small level of posting relates to, for example, patients blogging during their treatment.

Conclusion

The blogosphere may not have achieved the importance peer reviewed literature, but it is an area deserving of bibliometric research. Commercial blog search tools offer opportunities for this, but conclusions need to be drawn with care.

- Search tools can show trends in popular culture, particular if these are described by unique terms (e.g. Case Study 1)
- Interpretation requires awareness that words are being searched, rather than concepts, and national differences in language need to be taken into account (e.g. Case Study 2)

- Reference to a term (e.g. football) may be background to daily life rather than the subject of the blogging. This is a difference from searching for keywords in more focussed research material. (e.g. Case Study 2)
- There appears to be no easy way to track blogs relating to specific countries. (Case Study 3)
- References to institutions can be “polluted” by other uses of the words in the institution name (e.g. Case Study 4)
- It is important to sample blog postings to ensure that they are discussing topics being investigated. The prevalence of spam blogs in particular means that sudden recent spikes need to be treated with some scepticism. (e.g. Case Study 5).
- Scientific developments and controversies are reflected in blog discussions and this could be a useful way of evaluating the impact of science on society (e.g. Case Study 6)

BlogPulse is not designed for bibliometric purposes, so it is possible that dedicated search engines might reveal more about the blogosphere in future. Research blogs are appearing (e.g. Wikimetrics Research Blog, <http://wm.sieheauch.de/>) in which researchers offer commentary on their research area. In future a specialised search engine ("BlogScholar", perhaps, by analogy with GoogleScholar) would cover research blogs, providing early warning of research trends.

References

- Adar, E., & Adamic, L. A. (2005, Sept. 19-22, 2005). Tracking Information Epidemics in Blogspace. Paper presented at the *Web Intelligence 2005 Conference, Compiègne, France*. Retrieved November 23, 2006 from <http://www.hpl.hp.com/research/idl/papers/blogs2/>
- Auty, C. (2005). UK elected representatives and their weblogs: first impressions. *Aslib Proceedings*, 57(4), 338 - 355.
- Bar-Ilan, J. (2005). Information hub blogs. *Journal of Information Science*, 31(4), 297-307.
- Clyde, L. A. (2004). Library weblogs. *Library Management*, 25(4/5), 183 - 189.
- Cohen, E., & Krishnamurthy, B. (2006). A short walk in the Blogistan. *Computer Networks*, 50(5), 615-630.
- Delwiche, A. (2005). Agenda-setting, opinion leadership, and the world of Web logs. *First Monday*, 10(12). Retrieved November 23, 2006 from http://www.firstmonday.org/issues/issue10_12/delwiche/
- Dickey, M. D. (2004). The impact of web-logs (blogs) on student perceptions of isolation and alienation in a web-based distance-learning environment. *Open Learning: The Journal of Open and Distance Learning*, 19(3), 279-291.
- Glance, N. S., Hurst, M., & Tomokiyo, T. (2004). BlogPulse: Automated Trend Discovery for Weblogs. Paper presented at the *WWW2004, May 17-22, 2004, New York, NY* Retrieved November 23, 2006 from <http://www.blogpulse.com/papers/www2004glance.pdf>
- Herring, S. C., Scheidt, L. A., Wright, E., & Bonus, S. (2005). Weblogs as a bridging genre. *Information Technology & People*, 18 (2), 142-171.
- Jackson, N. (2006). Dipping their big toe into the blogosphere: The use of weblogs by the political parties in the 2005 general election. *Aslib Proceedings*, 58(3), 292-303.
- Leydesdorff, L., & Curran, M. (2000). Mapping University-Industry-Government Relations on the Internet: The Construction of Indicators for a Knowledge-Based Economy. *Cybermetrics*, 4(1), paper 2. Retrieved November 23, 2006 from <http://www.cindoc.csic.es/cybermetrics/articles/v4i1p2.html>
- Raynes-Goldie, K. (2004). Pulling sense out of today's informational chaos: LiveJournal as a site of knowledge creation and sharing. *First Monday*, 9(12). Retrieved November 23, 2006 from http://www.firstmonday.org/issues/issue9_12/raynes