Software measures temperature of blogosphere

Brian Ulicny

New technology allows us to anticipate and prepare for emerging trends in public opinion by analyzing activity in the blogosphere.

The blogosphere—a term that describes the ever-growing collection of Web logs, or blogs, on the Internet—provides a novel window into public opinion. Yet its dynamic, grassroots nature makes it an elusive medium to analyze and interpret in the aggregate, where its greatest potential lies. New technology is solving this problem by reading the signals of the blogosphere and zeroing in on issues that are most likely to migrate offline, enabling us to anticipate the opportunities (or threats) they present and to prepare for their impact.

There are nearly 16 million active blogs on the Internet, and more are being launched every day. Much of what is discussed in the blogosphere is fleeting and of little consequence to most of us. Increasingly, however, blogs are emerging as powerful organizing mechanisms, giving momentum to ideas that shape public opinion and influence behavior. The blogosphere can be a great bellwether of changing attitudes and new schools of thought, but only if we know which issues to pay attention to and how to identify those issues early in their life cycle.

Technology being developed at VIStology Inc., a research firm based in Framingham, MA, gives blog analysts a tool to monitor, evaluate, and anticipate the impact of blog content by clustering posts around news events and ranking their significance by relevance, timeliness, specificity, and credibility. These rankings provide useful analysis of the quality and significance of information in posts and are reliable and timely indicators of what is taking hold in the blogosphere.

Today, state-of-the-art blog search technologies allow the aggregation of posts only by the particular URL that they cite, not the event the post is about. By tethering the search function to a single data point, and a potentially misleading one at that, we seriously undermine our ability to recognize a topic’s importance and growing influence. Moreover, current technologies allow users to rank topics only by the popularity of a specific news article or by overall popularity of the blog itself. Techniques like these favor attention-grabbing posts that generate interest in the short term, but fail to appropriately highlight topics that reveal their significance only over time through a longer tail of interest and attention.

It is clear that traditional information retrieval techniques are ill suited to blog searching because of the unique nature and form of the medium. Blog posts are typically short, highly contextualized nuggets of information or commentary that depend on external links to fully convey their meaning. Traditional keyword searching is thwarted by this approach because major content elements, such as names of prominent participants in an event or specific examples supporting the writer’s thesis, may not even appear in the post itself. Without these elements, searches can easily bypass highly relevant information and leave valuable sources unidentified. Good blog searching requires indexing a blog’s references through links as well as through what is explicitly stated.

Search engines like Google became technically superior in Web searching because they rank Web documents not just by the frequency and position of query words in the document, but by the quality of the Web pages that link to the document. This is the essence of Google’s famous PageRank algorithm. Yet with blogs, individual posts are only rarely linked and require time to accrue the links they do attract. Search engines like VIStology’s are developing new approaches to assessing blog quality.

Of course, some blogs are significantly more popular than others as measured by the number of links they attract or the size of their audience. As with all content, however, popularity is not an effective measure of quality. In recent presentations at the International Conference on Weblogs and Social Media, and SPIE’s Defense and Security Symposium, I and my VIStology colleague Kenneth Baclawski presented statistics on tenured professors who have demonstrated through their academic careers that they are highly credible in their subject, yet have failed on average to attract a wide audience as bloggers. Conversely, popular group blogs may draw many readers and allow bloggers with marginal credibility to reach a much larger audience.

Continued on next page
In addition, we showed that more credible tenured professor bloggers can be objectively distinguished from average bloggers by means of certain features that convey trustworthiness, such as authoring content using their full names and citing reputable sources explicitly. These features, along with several others, form the basis of a credibility metric for blog posts that is part of VIStology’s technology. This is not to say that only tenured professors can achieve the highest level of credibility as bloggers; rather, tenured professors, as skilled conveyors of trustworthy information, demonstrably employ techniques that any other blogger can use to improve his or her reputation for veracity.

VIStology’s international blog-mining technology is being funded by the Distributed Intelligence Program of the US Air Force Office of Scientific Research under a three-year contract. This technology is being pursued as useful in its own right, and also as a test case in applying reasoning about information quality to a real-world application, using the network effect of the attention of the world’s bloggers to discover useful information in a timely way.

Author Information

Brian Ulicny
VIStology Inc.
Framingham, MA

References