

**Yijun Gao,**  
**Huron University College, The University of Western Ontario, London, ON**

## **Using the Web Server Logs to Study Public Opinion**

**Abstract:** With the help of Webometrics techniques, we could explore whether or not the Web surfer's online interest reflects the public opinion off-line. This paper investigates the Chinese Web user's interest regarding the United States and Japan, and demonstrates that Web server log data could be a good source for us to gauge the public opinion on specific domestic and international issues.

**Résumé :** À l'aide de techniques webométriques, nous avons pu déterminer si les intérêts des internautes reflétaient l'opinion publique hors ligne. Cette communication porte sur les intérêts des internautes chinois pour les États-Unis et le Japon et démontre que les données des journaux de serveurs web peuvent être utiles pour prendre le pouls de l'opinion publique sur certains enjeux nationaux et internationaux.

### **1. Introduction**

Analyzing the Web server logs offers us an unprecedented opportunity to investigate how readers access online news media. What we can learn from such usage patterns is of significance to both information scientists and journalists (Thelwall, Vaughan & Bjorneborn, 2005). The greatest advantage of the Web server logs is that such data show what people have actually done, and not what they say they might do or thought they did (Nicholas et al, 2004). Thus, information obtained from Web server log analysis is more objective and reliable than results of questionnaires or surveys.

Besides news reader's general online interest, we could conduct some studies of social phenomena, and explore whether or not the Web surfer's online interest could reflect the public opinion off-line, with the help of modern Webometrics techniques. This study investigated the Chinese Web user's interest regarding the United States and Japan, not only China's two biggest trade partners but also political rivals and economic competitors.

There are anti-American and anti-Japanese sentiments among average Chinese people nowadays due to historical and current causes. However, the Chinese government has been calling on its citizens looking to the future and decides to maintain close economic ties with the United States and Japan since the 1980s. Will the Chinese people's interest/opinion in the two countries be swayed accordingly?

The number of Web users in China, the world's largest, hit 384 million by the end of 2009, and more than 80.1% of these people rely on the Web to access news, according to the *25th Statistical Report on the Internet Development in China* (CNNIC, 2010). Such phenomena should not be ignored by the information science community.

### **2. Methodology**

Web server log statistics from the *People's Daily Online*, the biggest online newspaper in China, for one-year period (from October 2007 – September 2008), were collected and analyzed. Considering the sheer volume of the raw server log data, the author decided to use the statistical data generated by the WebTrends Log Analyzer installed at the server of the *People's Daily Online*, which efficiently helped us exclude visits generated by

“online noises” (such as Web crawlers' work, the graphics, banners, menu frames on the webpages, as well as malicious hits by the hackers or proxies).

The hypothesis for this study is that the Chinese online readers tend to access more “negative news reports” on the United States/Japan-related issues than those “non-negative stories”, in other words, “reports of human factor-involved negative events of the United States and Japan” attract more visits (pageviews).

The negative news reports on United States and Japan published by the *People's Daily Online* includes critiques/suspicious/worries about the two governments' policies, accidents/disasters involving human-factors, and natural disasters regarding them. Non-negative reports focus on cultural events, economic development, U.S. presidential elections, Sci-Tech progress, etc.

Because readers generally pay more attention to natural disasters (such as earthquakes) around the world, and there were not enough such news items (around 10 for the United States and 6 for Japan) during the data collection period, the "negative news reports" in this study were limited to reports on the "events involving human or governmental factors", such as plane crashes, traffic accidents, crimes, political turmoil, financial crisis, and attacks against the U.S. soldiers/citizens in Iraq, Afghanistan or other nations.

Considering online readers may actually like to access “negative news reports”, conclusions made from the U.S./Japanese related stories might be vulnerable. Therefore, we need a “contrast/control group” (i.e., negative/non-negative news reports on another country) to verify our findings from examining United States/Japan-related stories. Due to the international status of the United States and Japan, we find an influential player on the world stage---Russia.

Compared with the strong anti-American and anti-Japanese sentiments among the average Chinese people in the past two decades, there is no obvious anti-Russian sentiment in China. In the meantime, there are far fewer restrictions on reporting international events in China, which offers a good opportunity for us to investigate reader's interest in the United States and Japan. Investigating the public opinion of the two nations could make some contributions to the social phenomena studies.

Given that the WebTrends Log Analyzer for this project only provides the top 50 news reports' pageview numbers each day, the author first tried to collect one negative and one non-negative United States/Japan-related Chinese report from the same day's top 50 list. Once there were no such paired stories available from the same day, the author retrieved the needed pair within the shortest time period, with the restriction that the interval between the two stories should be no more than 10 days to reduce the possibility of visit fluctuations. Under these circumstances, when there was more than one report in a particular category, only the one with the highest pageview number was chosen.

The sample sizes for the two categories of news reports were kept equal: 30 negative and 30 non-negative stories, so as to ensure the power of the statistical tests. Two correspondents with the *People's Daily Online* were invited to re-classify the collected negative and non-negative U.S./Japanese-related stories respectively, with an aim of conducting an inter-coder consistency check and minimizing possible subjective factors. The inter-coder consistency was found to be very high (more than 95%).

The variance of pageview numbers is an indirect measure of the Chinese reader's interest in the U.S./Japanese-related issues, as well as the public mood. A more direct gauge of public sentiment would be the contents of comments posted to the news reports. This study further investigated reader's interest through a qualitative content analysis of the comments posted to the reports on the United States, Japan.

For each of the selected U.S./Japanese-related reports, the author randomly sampled four comments posted by the readers. Thus, there were 240 comments on the United States/Japan-related non-negative news reports and 240 comments on the United States/Japan-related negative news reports. Positive Comments are those supporting the United States or Japan; Negative Comments are those with anti-American/Japanese sentiments; Neutral Comments are all others objectively analyzed United States/Japan related issues, and were neutral in tone (neither positive nor negative).

One editor from the *People's Daily Online* helped the author classify the collected comments on the United States/Japan-related stories. The inter-coder consistency was also found to be high (more than 97%).

Collecting and classifying the Russia-related data (both news reports and comments) was straightforward: we just followed the rules set in the previous sections on the U.S./Japanese issues: 30 negative and 30 non-negative news reports on Russia-related events were collected, and each report also had four randomly sampled comments to be analyzed. The inter-coder consistency rates for the Russia-related reports and comments were also very high.

### **3. Results and Conclusions**

This study found that the Chinese news readers were more interested in reading negative news stories on the United States and Japan, since the negative reports attracted more readership as measured by pageview numbers. A similar phenomenon did not occur for Russia-related issues. This conclusion was supported by a qualitative content analysis of the comments posted to the United States/Japan/Russia-related news reports. Such findings reflected the public sentiments toward the two economic powers and Russia.

In other words, findings from the virtual community reflected the Chinese people's attitude in the real world. This demonstrated that Web server log data could be a good source for us to gauge the public opinion on specific domestic and international issues in future studies. However, putting too many U.S./Japanese-related negative reports on the homepage is not a very professional idea to attract readers, since such stories might mislead the public opinion and stimulate extreme nationalist sentiments in China.

This study, which reports findings from the author's Ph.D. thesis research, is related to the CAIS 2010 conference theme and addresses the third issues listed in the Call for Papers Poster (Human-Information Interaction---- user studies, information behavior).

### **4. References:**

CNNIC (2010). *The 25th Statistical Report on the Internet Development in China*, by Beijing: China Internet Network Information Center.

Nicholas, D., Huntington, P., Williams, P. & Dobrowolski, T. (2004). Re-appraising

information seeking behavior in a digital environment, *Journal of Documentation*, 60(1), 24-43.

Theilwall, M., Vaughan, L. and Bjorneborn, L. (2005), "Webometrics", in Cronin, B. (Ed.), *Annual Review of Information Science and Technology*, Information Today, Inc., Medford, NJ, 39, 81-135.