

Yijun Gao
Ph.D. Candidate
Faculty of Information and Media Studies
The University of Western Ontario
London, ON, N6A5B7
Fax: 519-6613506
Email: ygao3@uwo.ca

**From Editors' Choice to Readers' Favorites:
Analyzing Server Logs of China's Biggest Online Newspaper**

Abstract

Analyzing the Web server logs of the online newspapers will help us better understanding the readers' information needs. Web server logs from the People's Daily Online, the biggest online newspaper from China, were studied and revealed some important findings: The editor's choice of important news titles, which were put on the most prominent sections of the homepage, did not match with the top news ranked by their daily pageview numbers. In the meantime, this study also found that the Chinese and English readers' online interest in the same events were different.

Résumé

1. Introduction and related studies

The Web has become a full-fledged medium to disseminate news in the 21st century. Migrating newspapers to the Web will reduce or eliminate the harsh problems facing today's traditional newspaper industry, i.e. the falling circulation, lost of advertising revenues, soaring cost, delivery delays and waste generated by the printing systems. Some renowned English newspaper, such as the Seattle Post-Intelligencer and the Christian Science Monitor, even stopped daily publishing of paper-based editions in 2009 to survive the ongoing financial crisis (Ngowi 2009).

How Web surfers access online newspaper and what we can learn from such usage patterns are of significances for the information economy, journalism and Webometrics studies (Thelwall, Vaughan and Bjerneborn 2005), which can be accomplished through analyzing the Web server logs. The greatest advantage of the Web server logs is that they are records of what people have actually done, and not what they said they might do or thought they did (Nicholas et al 2004).

Information (i.e. readers' interest) obtained from Web server logs analysis is much more reliable and objective than results of answers from traditional surveys or questionnaires, and is crucial for the online newspaper's future developments, which highlights the significant transitions of traditional information institutions in the 21st century.

In addition to learning the general usage patterns of the online newspapers, i.e. which categories of news stories are popular among Web surfers, there is another interesting issue deserves study: whether or not the top news' various positions on the homepage of the online newspaper will yield significantly different pageview numbers. In other words, will the editor's choice of the important news match the readers' favorites?

The data source for this study is the People's Daily Online, the Web portal for the biggest national newspaper in China, where the number of Web users has swelled above 298 million by the end of 2008, after overtaking the United States in June 2008 to become the world's largest, according to the *23rd Statistical Report on the Internet Development in China* issued by China Internet Network Information Center (CNNIC 2009). The same report also noted that more than 234 million Chinese people rely on the Web as their major source of news, and online readers in China praised the People's Daily Online as the "most trustworthy news resource" (CASS 2002).

There are some previous news websites studies analyzing the server logs data or hyperlinks pointing to/going out of specific news sites. Among them, Nicholas et al (1999a, 1999b, 2000) investigated Web server logs of the Times and Sunday Times website from various perspectives, and explored new ways to study such valuable data. Wu and Bechtel (2002) studied the relationship between types of news events and daily traffic at the Web edition of the New York Times, and found that level of disruptiveness and episodicity were positively correlated with online traffic.

For the news website hyperlinks studies, Tremayne (2004) noted that the numbers of hyperlinks embedding in news stories on the web over a five-year period were increasing. Gao and Vaughan (2005) then established Web hyperlink profiles for four leading online newspapers. Tsui (2008) further pointed out that news articles from online editions of four major U.S. newspapers were lacking in the use of outgoing links pointing to resources beyond their domains. Besides, He and Zhu (2002) examined the development of China's online newspapers (including the People's Daily Online) in respect to the social environment framework, but did not mention their usage patterns. However, few studies focused on the top news homepage placement and their pageview numbers issues.

More recently, Bar-Ilan, et al. (2009) investigated user preferences for different rankings of search results from three major search engines (Google, Yahoo! And MSN Live), and noted that "placement on the page (i.e., whether the result appears near the top) is the most important factor used in determining the quality of the result, not the actual content displayed in the top-10 snippets" (Bar-Ilan et al. 2009, 135). Will similar phenomenon happen among the online news readers?

2. Methodology

First of all, we need to define “Editor’s Choice of Important News” and “Editor’s Use of ‘Best’ Homepage Positions”.

2.1 Editor’s Choice of Important News

At the People's Daily Online, there is one well known internal editorial rule: "put the most important reports on the most prominent places of the homepage to increase its visibility, and then to attract more readers", which is widely accepted by editors of this website. Thus, staff with the People's Daily Online's Chinese news channels not only updated the contents around the clock, but also tried their best to compete for the so-called better positions for the news from their own channels, since these employees have been evaluated by the total number of visits to the news/information from their channels.

However, another editorial rule implies that the news titles of less or no political / economic significance cannot be put on the most obvious/visible sections of the homepage. The "significant news" here refers to those on the Chinese leaders' activities or speeches, policies of the Chinese government, and the major international events. Sometimes, a few keynoted editorials of the People’s Daily or other opinions articles also enjoy such status and are published on the "most obvious position".

In addition to the Chinese reports, the People's Daily Online also translates and releases the “Editor’s choice of the most important Chinese news” into English and put them on the “best” parts of the English homepage, which offers us an unprecedented opportunity to explore the changing of online readers’ interest between different languages users on the same topics, and to provide more comprehensive answers to the research question of this study.

2.2 Editor’s Use of “Best” Homepage Positions

Opening the Chinese/English homepage of the People's Daily Online, all titles on the first screens are those on Chinese and other nations' political / economic issues, if the readers do not scroll down the page to browse more titles. Titles here refer to the news headlines on the homepage without related pictures. The homepage photos sections tell us another story: some much lighter topics, such as the beauty pageants, new discoveries around the world, and culture/sports events will show up there instead of the tedious top leaders "shaking hands ceremonies with foreign guests" (pictures).

For the People's Daily Online Chinese Edition, two sections on the homepage deemed to be the best: news title beneath the banner/Channel section (top of all other titles) and the homepage photos. Reports placed on the first position will have their font size twice as large comparing to any other titles listed on the homepage. Editors believe that the bigger font size titles and the photos will help top news published there attract more visits than the plain text, regular sized titles on all other sections of the homepage. Can editors’ hope become a reality?

For the People's Daily Online English Edition, the two sections for the most important news in the editors' minds are the Photo on the left and the Big Font Sized News Titles segments on the right at the upper half of the homepage. Under them are titles considered with "long term interest to the readers", which are generally left there un-changed for more than three days, until something more important appears. News offered in this section includes those on the Communist Party or China's top leaders' speeches and visits, or the Chinese government's policies. The daily updated news reports are laid at the lower half of the homepage in line with their publishing time (See snapshots of homepages in the Appendix 1).

2.3 Research Question

Since news titles with larger font size and the photo news will never show up again in other sections of the homepages, we could try to address this research question in the present paper: whether or not the news titles with the "larger font size", being put on "the upper half of the homepage" or the "homepage photos" rather than contents of the news titles, make the top stories attract more visits (pageview counts)?

Answering this research question will help us better understand the differences between "editor's choices" and "readers' needs" from the Chinese and English perspectives, which also reflect the online new readers' navigation behaviors.

2.4 Chinese vs English Readers

Answering the research question of this study requires us to clarify the "Chinese" and "English" readers as well. According to the relative geographic distributions of the IP addresses generating visits to the People's Daily Online (both the Chinese and English versions) from the twelve randomly selected dates, the majority of the IP addresses (approximately 85% on average) were limited to China (mainland). In the meantime, China (mainland) also hosted some 77% IP addresses that generated visits to the English news of the People's Daily Online. In addition to that, the second ranking country of the Chinese and English IP address distributions was the United States.

The IP addresses distributions for the Chinese version were quite straightforward: most of these readers are those Chinese speakers live in China, as well as the rising number of Chinese speakers/researchers in the English world (i.e. the U.S.). It is safe for us to assume that the vast majority of the English version readers are foreigners living either in China or abroad. The assumptions of these two reader-populations are based on the following reasons:

- (1) There are increasing amounts of foreign students/visitors/businessmen in China, who need something about China from the most authoritative channel;
- (2) The percentage (77%) of the IP addresses from China generating visits to the English version is less than their Chinese counterparts (85%);
- (3) It is very unlikely for the Chinese people to read news released from the English news website run by a Chinese news agency, or vice versa.

2.5 Data collection

Considering the sheer volume of the raw server logs data, the author decided to use the statistics 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” or any other “abnormal traffics” (such as Web crawlers' work, the graphics, banners, menu frames on the webpages, as well as malicious hits by the hackers or proxies). This method of filtering out abnormal traffics from server log statistics has been used in other studies (e.g. Xue 2004).

Twelve randomly selected days' Web server logs statistics data of the People's Daily Online were collected from October 2007 to September 2008, which include the daily pageview numbers of the top 50 Chinese and English stories, as well as the relative geographic distributions of the IP addresses generating visits to the Chinese and English versions during the twelve days. With the help of the Internet explore, the author manually saved the Chinese and English homepages four times on each of the selected dates (9am, 12pm, 3pm and 21pm, Beijing Time), in accordance with the updating frequencies of the People's Daily Online, then retrieved the top 50 news' positions information on the homepages.

According to the editorial rules of the People's Daily Online, its Chinese homepages will be divided into the following section: Homepage Pic, A section, B section, C section. The English homepages will be divided as well: Homepage Pic, A section, B section and C section. For both of the Chinese and English homepages, the C sections could not be seen in the first page until the readers scroll down. It is obvious that besides the Homepage Pic, all the other codings here have different meanings on the English and Chinese homepages. (See the attached printed out pages of the two homepage for details).

There are two methods to study the homepage positions influence on pageviews of the top news: the first one pools all position data together then run proper statistics test to compare the differences of the top news' pageview numbers. It is easy and feasible for us to conduct the twelve days data in this way. However, if there is much more data to be analyzed, let's say, 1000 days, such analysis method will not be a good idea and will be less efficient. Fortunately, we do have the “Plan B” to deal with such issue.

The top 50 news' homepage position data does not evenly distributed among the various groups, thus, we should always use the smallest number of news counts for each position as the sample size of that day. For example, if one Homepage Position only yields five pieces of news in the top 50 lists (which is the smallest among various positions), we should also choose the top five samples from other positions' candidates. This is the efficient method to ensure the power of the results with the help of equal-sized samples. Altogether, there were 116 Chinese and 138 English reports being analyzed to answer the research question of this project.

For each day's Chinese and English top 50 reports, the author also manually picked up those on the same events, since the English news of the People's Daily Online are translated from this online newspaper's Chinese reports. The pageview numbers of the selected titles were collected and saved in an Excel file. For each of the twelve randomly selected days, the author noticed that the average number of same story pairs is around 18 out of the 50 top Chinese/English news, which is good for us to conduct statistical tests to further compare the Chinese and English readers' interest.

3. Results and discussion

3.1 Top news homepage positions vs. pageview numbers

Since the frequency distributions of pageview counts of the sampled top 50 Chinese/English news are badly skewed, the non-parametric Kruskal-Wallis test was employed to study the differences of pageview counts among news with various positions. This test was carried out for both the Chinese and the English version of the of the People's Daily Online. Both tests revealed some interesting findings.

For the Chinese Homepage, the P-value is less than 0.01, which tells us that there are some significant differences among the pageview numbers of the four groups of news being compared. Although the test results do not specify the pattern of difference, judging from the median of pageview counts and average rank scores of the four groups, (see the third and fourth column of Table 1), we can see some pattern of differences.

Table 1: Kruskal-Wallis test for Chinese Top News on the Homepages

Home Positions for Chinese Top News	Number	Median of Pageview Counts	Mean Rank	P-Value
Homepage Pic	29	138470	40.79	0.00
A	29	257629	53.6	
B	29	531942	79.64	
C	29	299251	59.97	
Total	116			

It is safe to say that news putting on the "B" section attracted more visits than ANY other position because its average rank score is much higher than those of the other groups. However, there is no significant difference between the Homepage Pic and the A section, as well as A section and C section due to the fairly close average rank scores of their groups. Since the Mean Rank of the C section is higher than the A and Homepage Pic, we concluded that for the Chinese version of People's Daily Online, those titles put on the "least obvious parts" of the homepage will also have a good chance to attract more visits. It is clear that for the Chinese version of the People's Daily Online, editor's decisions of "importance" of news do not match the readers' interest.

There is a different story for the English homepage. According to Table 2, we could say that there is no significant differences among the visits to top news put at various positions of the English homepage. The average mean ranks as well as the medians of the

pageviews numbers of the three groups of news are relatively close. The mean rank difference between the Homepage Pic and the A section is not statistically significant, although the borderline P value of 0.122 is worth noting.

Table 2: Kruskal-Wallis test for English Top News on the Homepages

Home Positions for Chinese Top News	Number	Median of Pageview Counts	Mean Rank	P-Value
Homepage Pic	46	34885	63.3	0.122
A	46	68987	79.26	
C	46	48623	65.93	
Total	138			

Results of the two Kruskal-Wallis tests show that the Homepage Pic sections of the English and Chinese versions of the People's Daily Online do not attract more visits than plain text titles from other parts of the homepages. Ironically, the People's Daily Online's homepage photos cannot lure more visits as the editors expected them to do.

It is obvious that news titles put on the B section of the Chinese pages and the A section of the English pages both generated higher pageview counts than other positions. The findings suggest these two sections should be treated as the best sections on the homepage. The editors should put news titles with potentiality of being popular into this section. And the homepage photos section needs some revisions to play an even bigger role, because it may be true that some readers did not want to click the photo after viewing the mini ones on the homepages.

For the Chinese version, the A section (the title with biggest font size) does not attract more readers. However, there is little room for the People's Daily to change such thing due to the government's regulations, which requires only "the most important news in editor's mind" be placed on such position. However, the bigger font title on the English homepage (A section)'s performance is much better, though news titles put there are also stories of political or economical significance.

3.2 Chinese and English reports on the same events

There are 216 "paired" Chinese and English reports on the same events from the twelve days' data. Due to the pageview counts of the paired Chinese/English top news were also badly skewed, the Spearman's Rho test was applied to these paired stories on the same events, and the P value (0.41) for correlation coefficient is much larger than 0.05 level. We concluded that there was no significant relationship between the pageviews of Chinese and English news on the same events. Therefore, we could not estimate the English news ranking or polarity based on their Chinese counterparts' pageview numbers or vice versa. It is safe to say that the interest of Chinese and English are remarkable different with each other on the same events, while browsing the People's Daily Online.

The "paired" Chinese/English news study of this paper also found one interesting thing: the Chinese/English readers' interest on the same political event is dramatically different: the ranking of English news reports on the Chinese leaders' meeting with foreign dignities are generally much higher than their Chinese counterparts, though these two

kinds of news stories were of equal importance and put on the most prominent sections (i.e. A or B section of the Chinese homepage and A section of the English ones).

4. Significance of the study

This study addressed a classical question of information science: if and how users' information needs are being served through analyzing server logs statistics from a leading online newspaper in China, the biggest Web/information market of today's world. It examined and showed the usefulness of server log analysis in studying information needs. Results from this study show that the editors' choices and the readers' favorites do not always match each other, thus, content of news, i.e. its title on the homepage, is more important than its homepage position to attract more online visits, which is different with findings from Bar-Ilan et al (2009).

In the meantime, the Chinese and English readers' interest on the same events are also different: for example, many English reports of Chinese Top Leaders' activities received much more Web attention than their Chinese counterparts. Since the English and Chinese readers have different information needs, any bi-lingual or multi-lingual news website must tailor its content and homepage structures accordingly.

Based on the finding from this research investigation, the People's Daily Online has redesigned its English and Chinese homepages. For the English version, the "B" section used to be the place full of "the most popular news" from previous days, which were left there unchanged for five to seven days. The author noticed that these news will never be listed in the top 50 news again, which may waste the readers time to skip such "old information". The People's Daily Online re-designed its English homepage in 2009, and changed the "B"-section to a daily updated column. For the Chinese version, the size of the Homepage Pic was reduced during the homepage re-designing process. And, the editors there also began to pay much more attention to making the titles attractive than competing for the better positions on the homepage.

The study is interdisciplinary in that it relates to both information science and journalism studies. In such a globalization era, more and more people rely on the Web to retrieve news, methods and findings of this research are very helpful for those multi-lingual news websites in China or anywhere around the world (i.e. the U.S. or Canada's international news agencies) to thoroughly understand the online readers' interest, as well as the relative geographic coverage of specific online news media, so as to keep current customer and attract new users, which are all crucial for the information industry's sustainable developments in the coming years.

Acknowledgement

This paper reports findings from the author's PhD thesis research. Members of the supervisory committee: Professors Liwen Vaughan, Isola Ajiferuke, and Paul Benedetti, have been offering much valuable suggestion to this study.

Appendix: Homepage of the People's Daily Online



Figure 1: Homepage of the People's Daily Online (in Chinese)



Figure 2-2: Homepage of the People's Daily Online (in English)

References:

- Bar-Ilan, Judit, Kevin Keenoy, Mark Levene, and Eti Yaari. 2009. Presentation bias is significant in determining user preference for search results - A user study. *Journal of the American Society for Information Science and Technology*. 60(1): 135-149
- Chinese Academy of Social Sciences. 2002. *The National Survey of Web Use in China* Available at: <http://www.people.com.cn/GB/it/8219/29993/index.html> (in Chinese). Accessed January 6, 2009.
- CNNIC. 2009. *The 23rd Statistical Report on the Internet Development in China*, by Beijing: China Internet Network Information Center.
- Gao, Yijun and Liwen Vaughan. 2005. Web hyperlink profiles of news sites: A comparison of newspapers of USA, Canada, and China, *Aslib Proceedings: New Information Perspectives*, 57(5): 398 – 411.
- Ngowi, Rodrique. 2009. Christian Science Monitor has new look, new timing. *The Associated Press*. Available at: <http://www.msnbc.msn.com/id/29902792/> Retrieved: April 2, 2009
- Nicholas, David, Paul Huntington, Nat Lievesley, and Richard Withey. 1999a. Cracking the Code: Web Log Analysis. *Online and CD-ROM Review*, 23(5): 263-269.
- Nicholas, David, Paul Huntington, Peter Williams, Nat Lievesley, Tom Dobrowolski, and Richard Withey. 1999 b. Developing and testing methods to determine the use of web sites: case study newspapers, *Aslib Proceedings*, 51(5): 144-154.
- Nicholas, David, Paul Huntington, Nat Lievesley, and A. Wasti. 2000. Evaluating consumer Web site logs: case study The Times/Sunday Times Web site, *Journal of Information Science*, 26(6): 399-411.
- Nicholas, David, Paul Huntington, Peter Williams, and Tom Dobrowolski. 2004. Re-appraising information seeking behavior in a digital environment, *Journal of Documentation*, 60(1): 24-43.

Thelwall, Mike, Liwen Vaughan, and Lennart Bjorneborn. 2005. "Webometrics", in Cronin, B. (Ed.), *Annual Review of Information Science and Technology*, Medford, NJ: Information Today, Inc., 39: 81-135.

Tsui, Lokman. 2008. *The Hyperlinked Society: Questioning Connections in the Digital Age*. In Joseph Turow and Lokman Tsui. (Eds). *The Hyperlinked Society: Questioning Connections in the Digital Age*, by Ann Arbor: University of Michigan Press.

Wu, Denis H. and Arati Bechtel. 2002. Web site use and news topic and type *Journalism and Mass Communication Quarterly*, 79(1): 73-86.

Xue, Susan. 2004. Web usage statistics and Web site evaluation: a case study of a government publications library Web site, *Online Information Review*, 28(3): 180-190.