

**HOW ONLINE GATEKEEPERS GUARD OUR VIEW –  
NEWS PORTALS' INCLUSION AND RANKING OF MEDIA AND EVENTS**

CamLy Bui

PhD Candidate

Mass Communication Department

Newhouse School of Public Communications

Syracuse University

215 University Place, Syracuse NY 14228

[ncbui@syr.edu](mailto:ncbui@syr.edu)

### **About the author**

CamLy Bui is a PhD candidate from the Mass Communication Department of the Newhouse School of Public Communications, Syracuse University. Her research areas include political communication, global communication, and information and communication technologies. In particular, she is interested in exploring the extent to which new information and communication technologies affect information flow worldwide, and contribute to the formation and expression of public opinion in different socio-political systems.

## **Abstract**

This study examines two news portals, Google News and Yahoo News, using a theoretical framework that incorporates the theory of network gatekeeping (Barzilai-Nahon, 2007) and the notion of search engine bias. The study tests three hypotheses about the relationship between dominance of the news media, proximity of news events to the U.S. interests, and position of the news links on portals' front pages and result pages. The study analyzed 34,277 news items from 1,200 pages retrieved in 60 days in 2006 and 2008.

The descriptive data show four major trends: 1) Google News and Yahoo News differed significantly in media inclusion on the front pages, with Yahoo relying on a very limited number of media outlets; 2) the two portals also differed significantly in media inclusion on result pages except for 'Iraq bombing' pages in 2008: Yahoo relied more on agencies and U.S media outlets while Google relied more on non-major media from the U.S. and other countries; 3) both news portals increased the proportion of major media and decreased that of non-major media between 2006 and 2008; and 4) for both news portals, the distribution of media outlets was heavily skewed, with very few media outlets used hundreds of times during the studies period while the majority used only a few times. The hypothesis tests, using combined data of two portals of two years and separate data of each portal of each year, show none of the three hypotheses are fully supported.

These findings advance the understanding of the traditional gatekeeping notion in the Internet context. They also challenge the network gatekeeping theory regarding the

role of the gated relative to the gatekeeper, and caution against any sweeping generalization about news portals as a single entity.

**Keywords:** News portal, search engine, gatekeeping, online news, world news

## **Introduction**

Thanks to the thousands of news websites available online today, our view of the world has expanded in an unprecedented manner. The age of dependency on a few international news wires has long gone; we now enter an era where news from different corners of the world is constantly fed onto personal computer screens by both major global news media and numerous smaller news outlets. As with any type of online information, news on the World Wide Web is no longer scarce; it is now abundant.

Abundance, while clearly preferable to scarcity, has its own pitfall. Navigating through the intricate Web to get to a desired online destination is a daunting task, especially to inexperienced Web users. It is simply impossible for any individual to scan through all news websites, let alone thoroughly assess them and evaluate their credibility, hence the need for certain assistance. On the World Wide Web, news portals exist for such purpose, offering a broad array of worldwide resources and services to maximize the efficiency of online users who want to know what goes on around the globe.

The notion of news portals being the gateways to worldwide news is intriguing because it implies the process of gatekeeping. It is rather surprising, though, to find few scholarly works (e.g., Gerhart, 2004; Schroeder and Kralemann, 2005; Ulken, 2005) probing into the world of news portals. This study is an attempt to fill the void. It examines news portals, a type of web portals that provides aggregated news from thousands of news sites, often with the help of an embedded search engine for news search. The purpose is to understand how news portals, by way of news link display,

include and rank media outlets and news events on the front pages and the result pages. The world section pages of Google News and Yahoo News, the two most popular news portals in the U.S., are the investigated sites.

A recent survey shows approximately seventy five percent of Web users begin their search for online information through a search engine (Kennedy, 2005). According to the Pew Internet and American Life Project, the use of search engines ranks second only to email use as the most popular activity online ([Pew, 2009](#)). The popularity of search engines is also observed in users' quest for daily news: when it comes to news seeking, millions of users are turning to online news portals for their daily news fix (Lasica, 2004). In addition, research has shown that information seekers have the tendency to trust search engines' ability to rank results by their true relevance to the queries rather than making their own judgment about which results are in fact relevant (Pan, Hembrooke, Joachims, Lorigo, Gay, & Granka, 2007). In this regard, scholars even remark that they are the new gatekeepers between creators and consumers of online content (e.g., Hargittai, 2000). In light of the growing importance of these online tools in our daily life, it is high time to gain insight into their potential impact; this study is a timely effort in this regard. The study also provides new evidences regarding the online gatekeeping process and thus contributes to the overall advancement of the gatekeeping theory. Last but not least, the study offers a glimpse of the extent to which the Internet and the Web help expand – or hinder – our view of the world beyond geographic locality.

## **Theory**

The central argument that this study started out with is that web portals can allocate the attention of their users by acting as the gatekeepers to online information: The inclusion and ranking process makes certain pieces of information and sources more easily reached than others, and as the result, users are exposed to a limited package. Furthermore, in the specific context of news portals, the content of this limited package of news content is inherently biased towards U.S.-based mainstream media and U.S.-related news events. The reason for such bias is not necessarily due to any deliberate act, but because of the structure and technological root of the Web in general, and the hierarchical nature of the media worldwide. Two areas of recent research provide the framework and empirical evidences to support this argument: online gatekeeping, and bias on the Web.

### ***Online gatekeeping***

Since the first empirical study by D.M. White was published in 1950, gatekeeping has certainly been one of the most widely used constructs to investigate information dissemination through various communication channels. However, prior to the emergence of the Internet, and web portals in particular, gatekeeping in the traditional sense mainly dealt with the production process and ignored the role of the entities upon which gatekeeping was imposed (Barzilai-Nahon, 2007). Only at the turn of this century that the concept was revisited in response to the new media context. Hargittai (2000, 2003) and Introna and Nissenbaum (2000) are the first to mention the term ‘online gatekeeper’ when discussing gatekeeping implications of search engines and portals. Hargittai (2003) noted that the central concern “is no longer what is produced, but what consumers hear and know about” and that “gatekeeping activity still occurs online, but now takes place at the level of information exposure (p. 17).” A potentially strong

conceptual addition to the traditional gatekeeping notion, this new approach however has not been thoroughly developed in the research community until 2007, when Barzilai-Nahon proposed the theory of network gatekeeping.

The theory of network gatekeeping explores information control on the World Wide Web, positing that gatekeeping in a network context involves not only selection of information but also addition, channeling, manipulation, localization, integration, disregard and deletion of information. It defines a web portal as an *authority site*,<sup>1</sup> one of the network gatekeepers, and a search engine as a *channeling mechanism* in the gatekeeping process – a gateway station designed to direct information produced by the *gated* (i.e., the entity subject to gatekeeping) through the station. This theory proposes that the gateds, even though empowered with greater autonomy in the Internet era, are still largely dependent on the gatekeeper's design and policy to reach users due to the fact that attention of Internet users is concentrated on a very small number of information providers. The power of a gatekeeper over a gated is highest when the gated has little political power, less ability to produce information, minimal relationship with the gatekeeper, and fewer alternatives to circumvent gatekeeper's control; it is lowest when all four of these attributes are present on the side of the gated.

These propositions (Barzilai-Nahon, 2007) are applicable to this study of news portals for two reasons. First, they support the study's central argument that news portals do allocate users' attention towards some news media and events and not others. Second, following the argument of the relationship between the gatekeepers and the gateds, one can logically argue that big news agencies and mainstream media, especially U.S. mainstream media, have huge advantage over numerous other media around the world



because they have at least two of the four above-mentioned advantageous attributes, namely the ability to produce information and alternative mechanism to circumvent the gatekeepers' control.

### ***Bias on the Web***

A number of studies about portals and search engines have addressed the issue of information discrimination which, according to Barzilai-Nahon (2007), is the consequence of the online gatekeeping process. This volume of research calls information discrimination 'search engine bias,' which means the unequal treatment of websites and webpages in such a way that makes some websites more readily within reach of information seekers than others (Hargittai, 2000). Various studies in the field of computer science and information studies seem to point to the fact that, at the moment, the World Wide Web is structured and organized such that more prominent websites always get more new links than small sites (Kleinberg & Lawrence, 2001). The distribution of links in this manner follows what scholars define as power law distribution, which leads to what is referred to as the rich-get-richer phenomenon (Introna & Nissenbaum, 2000). In this environment, these scholars argue, search engines' preference of more prominent media seems to be unavoidable, the reason being that one of the major criteria for search engines' ranking is hyperlinks. When it comes to the news media environment, such bias means that a limited number of big media are more likely to be included and ranked in high order by search engines. The more links a website has and the more prestigious those links are,<sup>ii</sup> the higher it is in ranking order. Consequently, in theory, there is little chance for local, newly-formed or lesser-known media to compete with mainstream media.

This kind of bias, to computer science researchers, is primarily a technical issue and not a content-type bias (Couvring, 2005). Social science researchers, however, point out that such bias does result in unequal presentation of information, which clearly has certain societal implications. A study by Hindman and his associates (2003) reveals that, when it comes to political content on the Web, both Google and Yahoo allocate users' attention toward a remarkably small number of popular political websites. The scholars define such bias as "Googlearchy," and conclude that the hierarchy of traditional media does extend to the online world. The implication of such bias in politics, according to Hindman et al. (2003), is that political conversations take place on very limited number of websites leading to more concentration and polarization of ideas; moreover, Googlearchy presents structural barriers to democratic deliberation of average citizens rather than making all voices equally heard.

Empirical data support both believers and skeptics of the existence of online bias. Some studies (Vaughan & Zhang, 2007; Vaughan & Thelwall, 2004, Kleinberg & Lawrence, 2001) confirm that popular sites receive highest ranking in search results, while others such as Fortunato, Flammini, Menczer and Vespinari (2006) conclude that search engines are more inclusive than they are given credit. Two recent studies also provide contradicting results regarding news portals' bias. Lee's study in 2005 on the effect of search engines' news services on the online news market shows that: 1) all engines most frequently select news agencies as sources of information; 2) all engines focus on events of the Middle East and Asia; and 3) many news media included by the engines in fact use reports of news agencies which means engines get agencies' reports indirectly. However, Schroeder and Kraleman (2005), based on a survey conducted in

2004 comparing Google News Germany and mainstream German media, conclude that influential media groups do not set the agenda for Google News – on the contrary, Google News includes numerous news media from various countries, allowing for a broader range of news events.

Both studies, even though providing intriguing findings, were rather narrow in scope. Lee's study only used descriptive data to compare the number of news agencies used in portals; mainstream media, both from U.S. and other countries, were not taken into account. Schroeder and Kralemann's study, on the other hand, only limited their examination to Google News in Germany and its inclusion of mainstream German media.

### ***Research question and hypotheses***

At this point, it is important to re-emphasize that the news portals as gatekeepers exert but another level, and probably the last level, of gatekeeping. A search engine of a portal, while having control over what goes into the front page and the result pages, does not have complete control or in any way manipulate the content it finds. News items presented on the portal's pages are not created by the engine itself but by the media outlets from which it retrieves information. Various decisions have been made by other mass media prior to the arrival of a news report at the portal's index. Millions of news events are filtered by thousands of mass media, resulting in news menus presented in their respective websites. A search engine in turn filters this already narrowed pool of news to build its own index of significant events worldwide to feed onto the portal. The content of an online portal's news pages is thus bound by the content of the online news media worldwide.

The issue at hand is that different media make different decisions about news, arriving at different news menus. By including some media outlets while excluding others, a search engine consequently includes certain news events while excluding others. If a search engine picks up news only from the mainstream media, the news index on the portal will naturally be reflective of the news agenda of those media. On a practical level, it is certainly difficult for news portals to avoid being over-reliant on major media worldwide and to become more inclusive in their use of media outlets. Research by Boyd-Barrett (1997), Boyd-Barrett and Rantanen (1998), and McPhail (2002) arrived at a convergent point regarding media domination: international news agencies and global media giants (primarily from the U.S. and other Western countries) continue to rule the global media landscape, and the trend will stay unchanged in the foreseeable future. In other words, big media still dominate the news agenda worldwide, and local media are still strongly influenced by their agenda. Similarly, Derbyshire's article in 2005 shows that the five major media corporations – Time Warner, Disney, Bertelsmann, Viacom, and News Corporation – still rule in the global news market up to that moment. If big media's domination is still a reality, it would be difficult to expect search engines to pick up less prominent news media – after all, the sheer number of news reports published by agencies and mainstream media means a higher probability of inclusion in an engine's crawl.

Taking the cues from the above-discussed issues, this study put forth one research question and three hypotheses. The research question examined the overall inclusion of news media and news events on the front pages and on the result pages of Google News and Yahoo News, and would be addressed using descriptive data.

**Research question:** In terms of media breakdown, how prominent are worldwide news agencies and major news media compared to other news media on the front pages and result pages?

In addition, based on the above arguments and evidences about the dynamics of news media and events on news portals, the following hypotheses about news portals' ranking were tested.

**Hypothesis 1.** The more dominant the media outlet is in the online media system worldwide, the higher its news link is on the front page.

This hypothesis assumed a positive relationship between the dominance of the media outlet and the position of the news link on the front page. Variable dominance of the media outlet in the online media system worldwide was defined in this study as the extent to which the media outlet has high popularity, large readership, or strong influence over other news media in the media system. Variable position of the news link was defined as the relative rank of a news link on a front page or result page of the world section of news portals, with top position having the highest rank.

International agencies are often assumed to have the highest influence because they feed reports to numerous media outlets worldwide (Boyd-Barrett and Rantanen, 1998; Lee, 2005). It is logical to assume that major media in the U.S. are more likely to have higher readership than major media in countries such as Kenya because the U.S. media established their existence on the web earlier than most media in other countries did (Vaughan & Zhang, 2007). By the same token, major media are also often assumed to have higher popularity than non-major media such as news blogs or local news sites. If, as various accounts argued (see above), news portals follow the rule of bias, then it is

logical to assume that international agencies and major media (especially those from the U.S.) are ranked higher in position than other media in the online world.

***Hypothesis 2.*** A news event closer to the U.S. interest is more likely to be ranked in higher position on the front page than a news event further to the U.S. interest.

This hypothesis assumed a positive relationship between the proximity of an event to the U.S. interest and position of the news link. Variable proximity of an event to the U.S. interest was defined as the extent to which a news event is a U.S. event, a bilateral event that involves the U.S., a multilateral event that involves the U.S., or other event that does not involve the U.S. at all. Regarding news events on the front page, if it is true that more U.S. news media are included and ranked in high position as some scholars (e.g., Vaughan & Zhang, 2007, Lee, 2005) have found, it is quite logical to expect more U.S. or U.S.-related news events to be presented in higher position on the pages than other countries' news events.

***Hypothesis 3.*** Proximity of an event to the U.S. interest can mediate the relationship between the dominance of the media and the position of the news link. With an event closer to the U.S. interest, a more dominant media is likely to be in higher position on the result page; with an event farther from the U.S. interest, a less dominant media is.

Previous studies on search engines' rankings did not take into account the nature of searched news events and how it might make a difference in the ranking process. This study suspects that if the search term is about a news event closer to the U.S. interest, the result pages will privilege dominant media outlets over other news media, the reason being that more dominant news media especially those from the U.S. are more likely to

cover such event in greater length and depth and thus are deemed more relevant to the search algorithm. If the search term is about a news event farther from the U.S. interest, news portals will privilege media outlets other than the dominant ones on the result pages, the reason being that dominant media are less likely to cover such event than local media outlets so the latter have a better chance of being included. In this sense, whether an event is closer to the U.S. interest or not can be considered a mediating factor in media ranking.

## **Method**

The front pages of the world section and the first result pages of two search queries – ‘Iraq bombing’ and ‘Thailand trade’ – of both Google News and Yahoo News portals comprised the sampling frames of this study. Two thirty-day periods in April 2006 and January-February 2008, were chosen as the study’s timeframe. The unit of analysis is the news link, defined in this study as an item on the news portal that includes the headline of the news article by a media outlet, the hyperlink to the original media site, and in some cases a lead-in to the news report itself. A computer program was created for page retrieval. The pages were downloaded at a 12-hour interval (at 9:00 am and 9:00 pm).<sup>iii</sup>

In total, 240 front pages were retrieved resulting in 21,465 news links; 960 result pages were retrieved resulting in 12,812 news links. Since the front pages of Google News and Yahoo News are different in design, for comparative purpose, only the top five news links of the pages were sampled for hypothesis tests of ranking. In total, 1,200 news

links were included in the sample. In addition, five random links on the first result pages were used for hypothesis tests. In total, 2,400 news links were included.

### ***Independent variables***

Two independent variables were investigated: *dominance of the media in worldwide online media system*, and *proximity of an event to the U.S. interest*. A coding scheme was developed for variable dominance of the media such that a U.S. news agency was coded as 6 (assuming highest dominance value); non-U.S. agency as 5; U.S. major media as 4, other country's major media as 3, U.S. non-major media as 2, and other non-major media as 1. Country of the media was also coded, with 5 representing U.S. media, 4 Britain, 3 France, 2 Canada, and 1 other country. A list of major national news media, provided by the online service Mondo Times ([www.mondotimes.com](http://www.mondotimes.com)), was used as a reference list for mainstream media and news agencies worldwide.

For variable proximity of an event to the U.S. interest, a coding scheme was developed such that a U.S. news event was coded as 5 (assumedly closest proximity); bilateral event involving the U.S. as 4, multilateral event involving the U.S. as 3, multilateral or bilateral event not involving the U.S. as 2, and non-U.S. single-country event as 1. News links on the result pages of the search term 'Iraq bombing' were coded as 4, under the assumption that they represented a bilateral event involving the U.S., and news links on the result pages of the search term 'Thailand trade' were coded as 2, under the assumption that they represented a non-U.S. bilateral/multilateral event.

### ***Dependent variable***

One independent variable was investigated: *position of the news link*, defined as the relative rank of the link on either the front page or the result page. On all pages, the



ranks were set in a descending order (meaning links higher to the top of the page as seen on the computer screen were higher in ranking). For the sampled top five links on the front page, the link at the top was coded at 5 (highest value in rank). The one below it was coded as 4, followed by 3, 2, and 1. For the sampled five links on the result page, the news link in highest position was coded as 5, the lower ones as 4, 3, 2, and 1.

### ***Intercoder reliability***

A trained coder along with the researcher coded 10% of the sample (N = 3,600) to establish the intercoder reliability, then the researcher coded the rest of the sample. Cohen's Kappa test yielded a  $\kappa$  of .86 for variable media dominance, and .76 for variable proximity of an event to the U.S. interest. Pearson's  $r$  test yielded perfect agreement score for the dependent variable.

## **Results**

### ***Research question***<sup>iv</sup>

As Table 1a shows, in both 2006 and 2008, the total number of distinct media outlets on Google's pages far exceeded those of Yahoo's. Yahoo in fact relied on less than a handful of media. Google, on the other hand, uses news reports from dozens of news agencies and major media, and from hundreds of non-major media worldwide. However, 2008 saw a relatively significant drop in the number of distinct non-major media outlets (from 519 in 2006 to 331 in 2008) on the front pages of Google News. Even though both media categories see a decrease in total number of distinct media outlets in 2008, the proportion of distinct major media outlets actually went up against the proportion of distinct non-major media outlets (13% versus 87% compared to 9% versus 91% in 2006).

**Table 1a.** Total distinct media outlets on front page by portal

		<i>Major media</i>	<i>Non-major media</i>	
Google				
	2006	53 (9%)	519 (91%)	N = 572 (100%)
	2008	49 (13%)	331 (87%)	N = 380 (100%)
Yahoo				
	2006	3 (50%)	3 (50%)	N = 6 (100%)
	2008	4 (67%)	2 (33%)	N = 6 (100%)

Table 1b further reveals the apparently uneven treatment of media outlets. Yahoo News was highly reliant on major media outlets; in 2008, non-major media outlet on the front page of Yahoo News accounted for much less than 1% of total media inclusion. (A closer look into the media list of Yahoo shows that, within the major media category, AP appeared to be a dominant outlet with frequency of inclusion accounting for 57% of all major media outlets in 2006 and 55% in 2008). Google, although having a much larger variety of media outlets, also shows an inclination to increase its inclusion of major media outlets: In 2006, the percentage of major media outlets on Google's front pages was 46%; in 2008, it went up to 66%. A closer look at Google's media outlets in 2006

reveals that, within the non-major media category, media outlets from the U.S. and Great Britain were used much more frequently than media outlets from other countries.

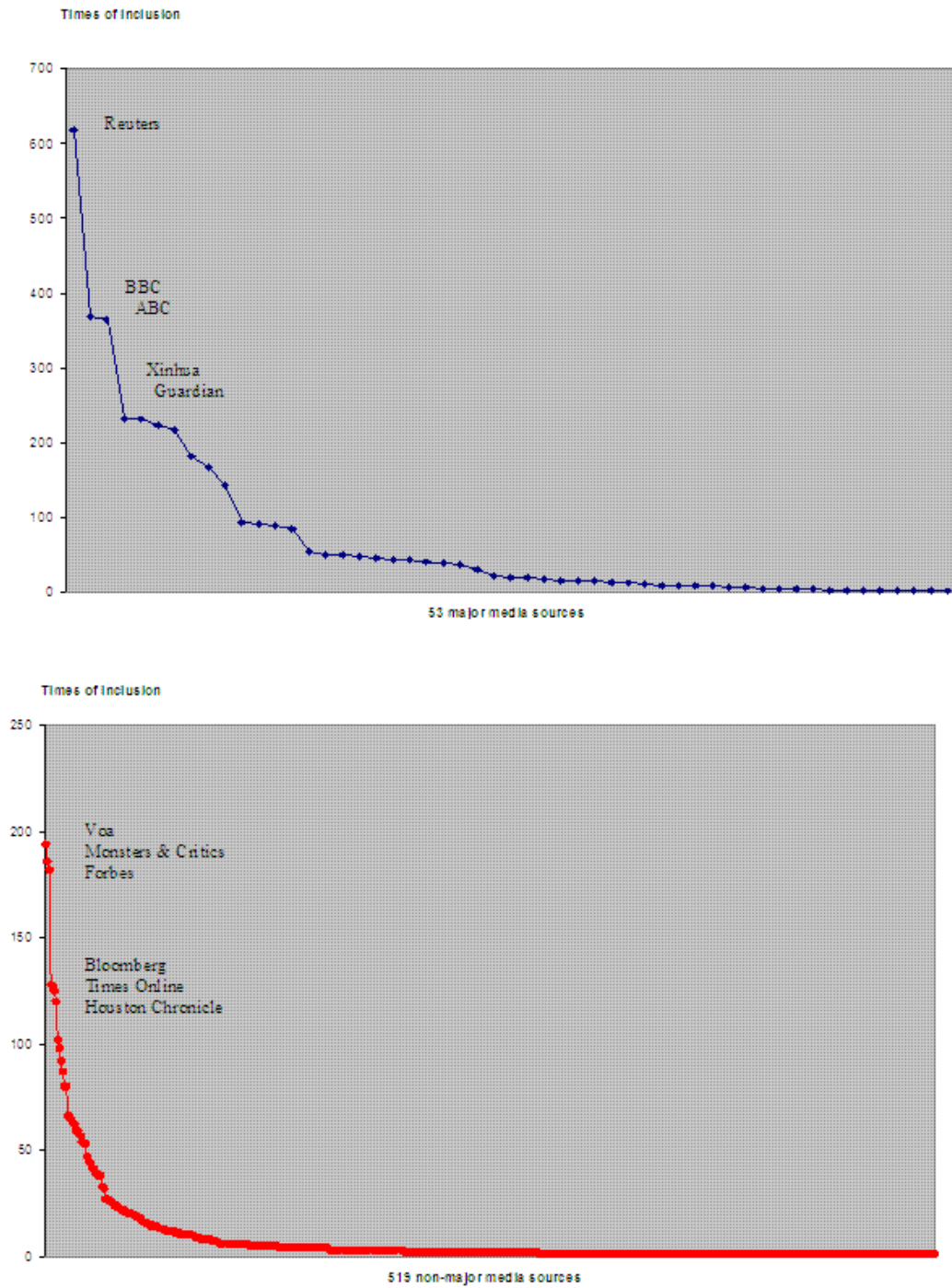
**Table 1b.** Media outlets by frequency of inclusion on front page by portal

		<i>Major media</i>	<i>Non-major media</i>	
Google	2006	3825	4488	N = 8313
		(46%)	(54%)	(100%)
	2008	5542	2810	N = 8352
		(66%)	(34%)	(100%)
Yahoo	2006	2290	110	N = 2400
		(95%)	(5%)	(100%)
	2008	2396	4	N = 2400
		(100%)	(0%)	(100%)

Also, there appears to be an increase in concentration on a limited number of media outlets on Google's front pages over the studied period. Figure 1 shows the distribution of distinctive media outlets on Google's front pages in 2006, which is similar in trend to the distribution in 2008. In the non-major media category, the top ten media outlets accounted for 30% of all non-major media outlets' inclusion in 2006; the percentage was 61% in 2008. Note that this percentage increased remarkably even when the total non-major media outlets decreased within the same time frame. Similarly, within the major media category, the top ten media outlets accounted for 72% of all major media outlets' inclusion in 2006; the percentage was 78% in 2008.



**Figure 1.** Distribution of distinct media outlets on front page of Google News in 2006



The difference between Google and Yahoo in media inclusion is not as remarkable on the result pages as it is on the front pages. As Table 2a shows, both portals relied on more non-major than major media outlets, although in terms of total number of distinct media, Yahoo still used fewer media outlets than Google did in both media categories. Moreover, both portals included significantly more media outlets on ‘Iraq bombing’ pages than ‘Thailand trade’ pages. In terms of media breakdown over the two-year period, Google slightly increased its inclusion of major media outlets and significantly decreased its inclusion of non major media outlets (from 371 to 238 with ‘Iraq bombing’ result pages, 216 to 153 with ‘Thailand trade’); Yahoo had similar trend with ‘Iraq bombing’ result pages but a reversed trend with ‘Thailand trade’ pages.

**Table 2a.** Total distinct media outlets on result page by portal

		<i>Iraq bombing</i>			<i>Thailand trade</i>		
		<i>Major</i>	<i>Non-major</i>		<i>Major</i>	<i>Non-major</i>	
Google	2006	40	371	N = 411	28	216	N = 244
		10%	90%	(100%)	11%	89%	(100%)
	2008	42	238	N = 280	30	153	N = 183
		15%	85%	(100%)	16%	84%	(100%)
Yahoo	2006	26	145	N = 171	25	86	N = 111
		15%	85%	(100%)	23%	77%	(100%)
	2008	28	123	N = 151	18	98	N = 116
		19%	81%	(100%)	16%	84%	(100%)

Table 2b provides further information about the shift in media breakdown by frequency of inclusion. Between April 2004 and February 2008, Google appeared to increase its percentage of major media outlets and decreased its percentage of non-major media outlets; on ‘Thailand trade’ pages in 2008, major media outlets even became the majority. In the case of Yahoo, major media outlets also increased in percentage, and nearly closed the gap with non-major media outlets (from 41-to-59% in 2006 to 49-to-51% in 2008) on the ‘Iraq bombing’ pages. Note that major media percentage increased on ‘Thailand trade’ page even though the number of major media outlets dropped from 25 in 2006 to 18 in 2008.

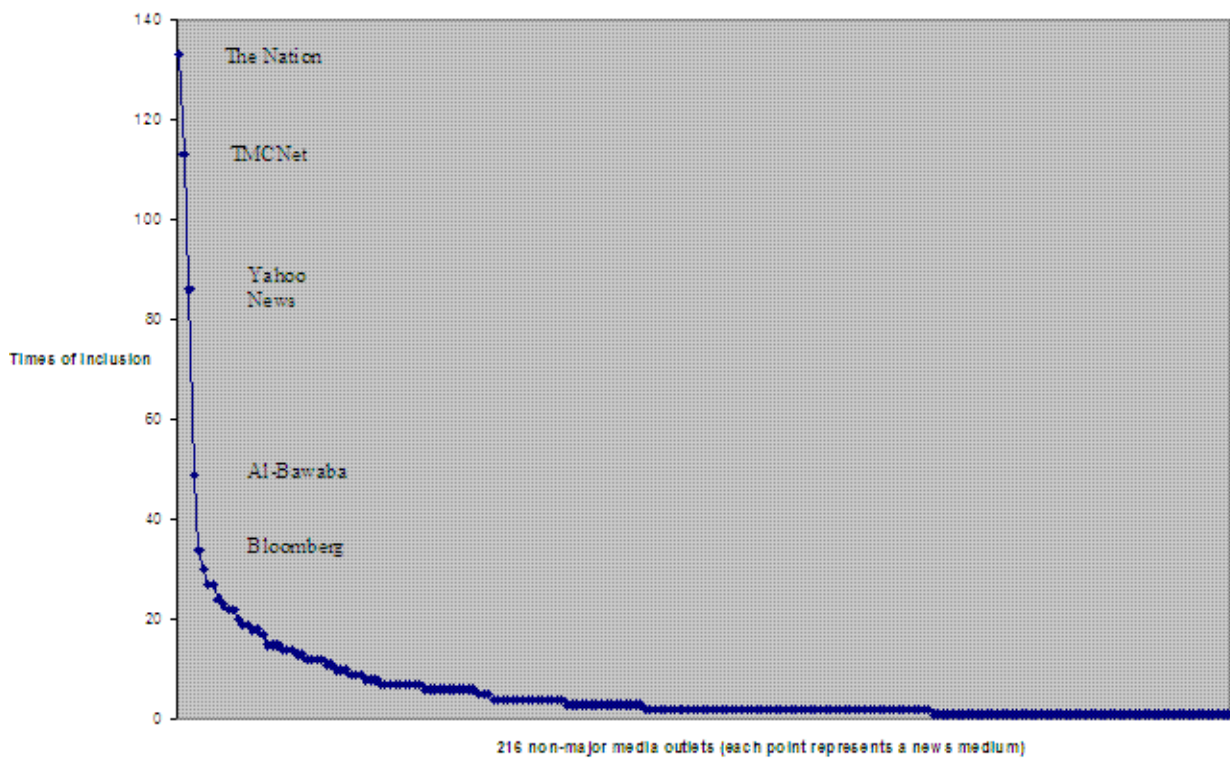
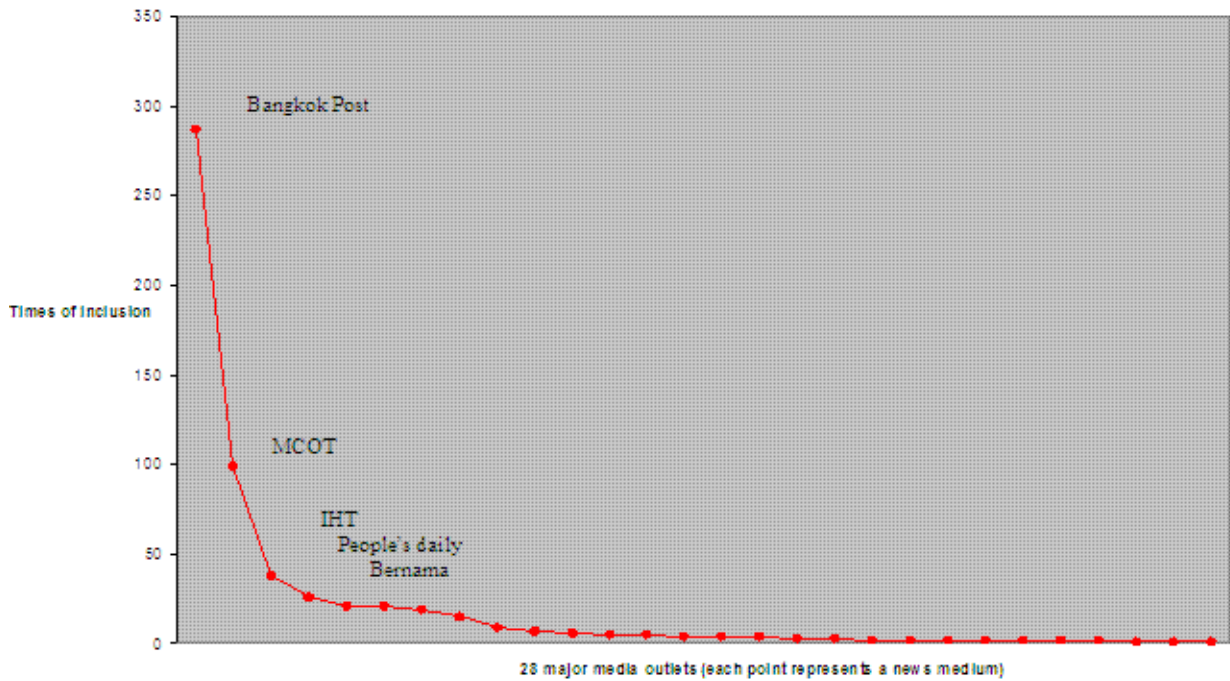
**Table 2b.** Media outlets by frequency of inclusion on result page by portal

		<i>Iraq bombing</i>			<i>Thailand trade</i>		
		<i>Major</i>	<i>Non-major</i>		<i>Major</i>	<i>Non-major</i>	
Google							
	2006	534 (23%)	1799 (77%)	N = 2333 (100%)	593 (30%)	1399 (70%)	N = 1992 (100%)
	2008	957 (43%)	1272 (57%)	N = 2229 (100%)	891 (56%)	706 (44%)	N = 1597 (100%)
Yahoo							
	2006	490 (41%)	710 (59%)	N = 1200 (100%)	471 (33%)	729 (61%)	N = 1200 (100%)
	2008	583 (49%)	617 (51%)	N = 1200 (100%)	520 (43%)	680 (57%)	N = 1200 (100%)

Within the major media category, there is a clear tendency to use the majority of news reports from a limited number of media outlets. The top ten most-used major media outlets accounted for 91% and 92% of all news links on the ‘Thailand trade’ result pages of Google in 2006 and 2008, and 74% and 73% of all links on the ‘Iraq bombing’ pages. For Yahoo, its percentages were 88% and 93%, and 87% and 79%, in the same respective order. Within the non-major media category, none of the media outlets was dominant, and the top ten most-used media outlets accounted for one-third or less of all news links on result pages of both news portals (except for the case of ‘Thailand trade’ pages of Yahoo, when top ten media outlets accounted for 48% and 54% of the links in 2006 and 2008). Figure 4 shows the distribution of distinct media outlets on ‘Thailand trade’ result pages of Google News in 2006, which is similar in trend to ‘Iraq bombing’ result pages of Google in 2008, and similar to both pages in Yahoo.



**Figure 4.** Distribution of distinct media outlets on 'Thailand trade' result pages of Google News in 2006



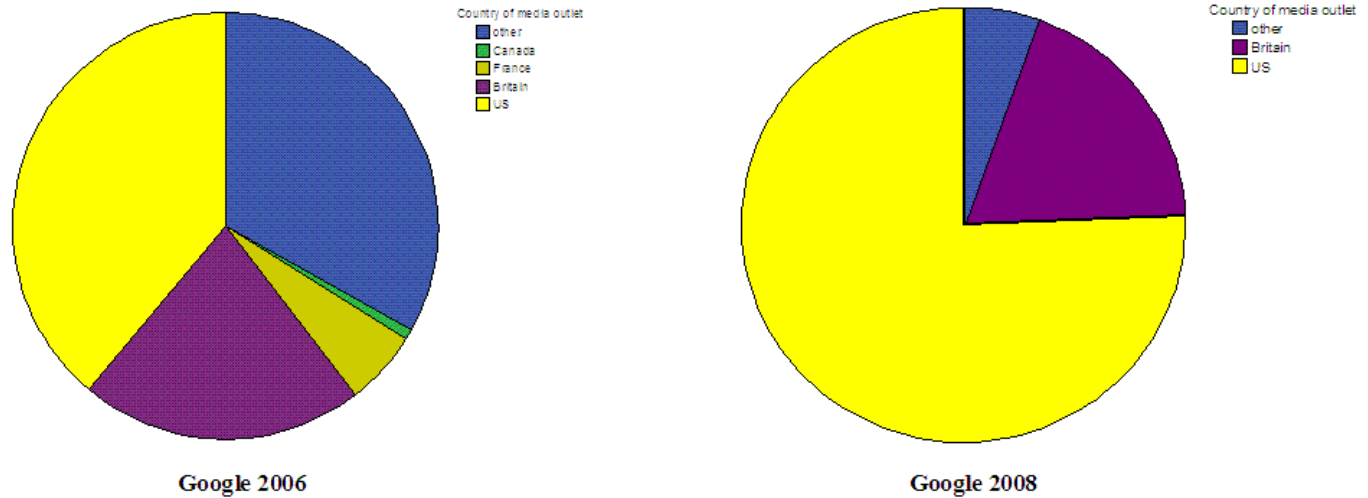
Looking specifically at the top five news links on the front pages, there is also a stark contrast between Google News and Yahoo News in media outlet inclusion as shown in Table 3. In 2006, Yahoo News relied on only two media outlets: one U.S. news agency (The Associated Press, 98% of the time), and occasionally one U.S. major newspaper (Christian Science Monitor, 2%). In 2008, only Associated Press' reports were used in the top five positions of the front page, making the agency the sole provider of top stories for Yahoo News.

**Table 3.** Percentages of media outlets in the top five position on the front page by portal

		<i>2006</i>	<i>2008</i>
Google			
	U.S. agencies	1.0	4.7
	Other agencies	7.3	12.3
	U.S. major media	16.7	34.7
	Other major media	28.7	23.7
	U.S. non-major media	31.0	23.3
	Other non-major media	15.3	1.3
		100%	100%
		(N = 300)	(N = 300)
Yahoo			
	U.S. agencies	98.0	100.0
	Other agencies	0	
	U.S. major media	2.0	
	Other major media	0	
	U.S. non-major media		
	Other non-major media	0	
		100%	100%

Google News appeared to include a wider variety of distinct media outlets in the top five positions, but the trend in 2008 differed significantly from that of 2006 ( $\chi^2 = 69.83$ ,  $df = 5$ ,  $p < .001$ ). A closer look at Google's proportion of media outlets again shows a clear shift toward more concentration of media outlets. In 2006, 85 distinct media outlets were included in the top five positions; the number dropped to only 30 media outlets in 2008. In 2006, about 46% of news reports in the top five positions came from non-major media outlets from the U.S. and other countries; in 2008, the number dropped to less than 25%. Moreover, within this non-major media group, non-U.S. media only accounted for a modest percentage – 15.3% in 2006, and only 1.3% in 2008. Meanwhile, more U.S. news agencies, other agencies, and U.S. major media were included in the top five positions in 2008 than they were in 2006 – their percentage all increased remarkably – as the percentage of non-U.S. major media slightly decreased within the same time frame. Overall, regardless of whether they are news agencies, major or non-major media, the majority of media outlets included in the top five positions of Google's front pages were from the U.S. or Great Britain (See Figure 2).

Figure 2. Proportion of Google's top five media outlets by country



In terms of event coverage, one would expect Yahoo to have a higher proportion of U.S. and U.S. related news on its front pages than Google since the portal relies solely on reports from AP and Christian Science Monitor. A chi-square test shows that there was indeed a significant difference between the two portals in event coverage in both 2006 ( $\chi^2 = 16.60$ ,  $df = 4$ ,  $p < .01$ ) and 2008 ( $\chi^2 = 10.10$ ,  $df = 4$ ,  $p < .05$ ). However, descriptive data show that Yahoo did not always privilege U.S. and U.S.-related news. As shown in Table 4, both news portals carried more non-U.S. events than they did U.S. and U.S.-related events in the top five positions of their front pages in 2006 and 2008. Even though Yahoo did carry more U.S. and U.S.-related events than Google in 2006, the difference was not really remarkable (32.3% versus 27.1%); in 2008 the breakdown changed somewhat (23.8% versus 28%).

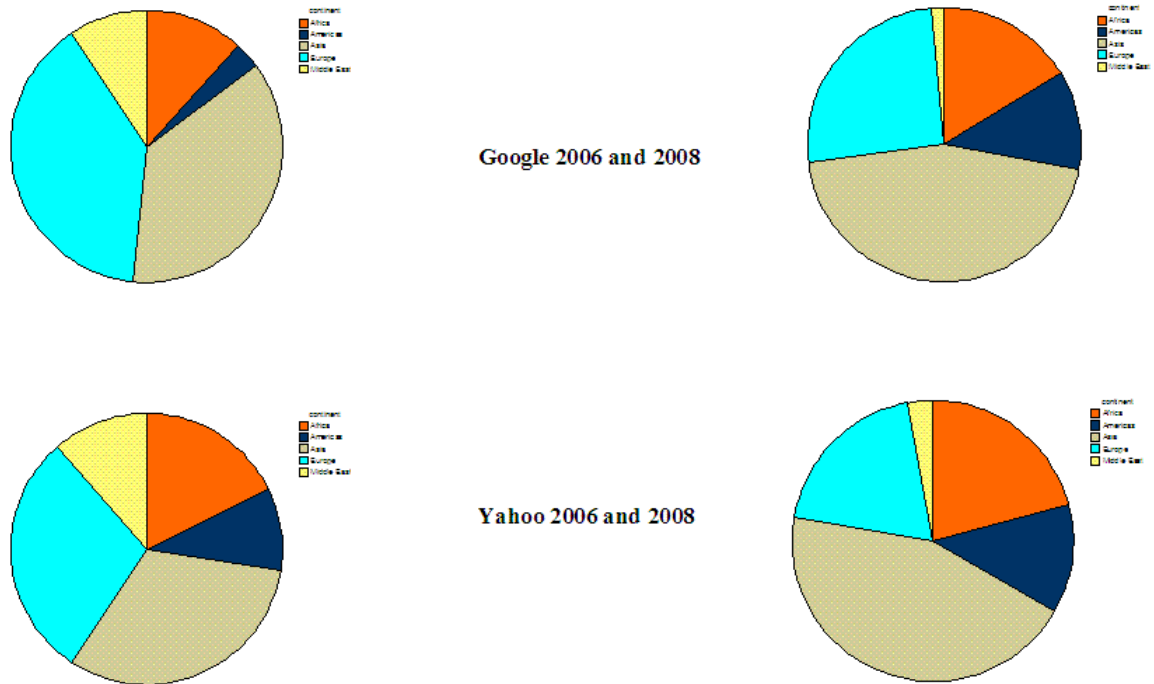
**Table 4.** Percentages of news event in the top five position on the front page by portal

		<i>2006</i>	<i>2008</i>
<b>Google</b>			
	U.S. event	4.7	1.7
	Bilateral event involving U.S.	6.7	8.0
	Multilateral event involving U.S.	15.7	18.3
	Other bilateral/multilateral event	22.3	26.7
	Single-country event	50.7	45.3
		100%	100%
		(N = 300)	(N = 300)
<b>Yahoo</b>			
	U.S. event	1.3	.3
	Bilateral event involving U.S.	13.7	10.4
	Multilateral event involving U.S.	17.3	13.1
	Other bilateral/multilateral event	15.7	21.8
	Single-country event	52.0	54.4
		100%	100%
		(N = 300)	(N = 300)

For both news portals, single-country events outside the U.S. accounted for the largest percentages in 2006 and 2008, which means more single-country events were present in the top five positions than U.S. events and bilateral/multilateral events. Interestingly, Iraq stories (not involving the U.S.) did not take up a large percentage within the single-country group, and Yahoo News appeared to pay more attention to Iraq than Google did (Google: 15.2% in 2006, 5.1% in 2008; Yahoo: 27.1% in 2006, 16.7% in 2008). With Iraq taken out of the picture, the majority of single-country news reports were about Asian or European countries (See Figure 3). Reports about African countries were also

given higher attention than those about countries in the Middle East or Americas (not counting the U.S.).

Figure 3. Single-country news event by continent by portal



With regard to the selected five news items on the result pages of the two search terms, Table 5 shows the breakdown of media inclusion into six categories: U.S. news agencies, other news agencies, U.S. major media, other major media, U.S. non-major media, and other non-major media. Chi-square test shows that the difference in media inclusion on result pages between Google News and Yahoo News in 2006 and 2008 is statistically significant when the search term is ‘Thailand trade’ ( $\chi^2 = 57.80$ ,  $df = 5$ ,  $p < .001$  and  $\chi^2 = 46.95$ ,  $df = 5$ ,  $p < .001$ , respectively). Comparing within each media

category, Google in 2006 used more non-major media (both U.S. and other) and non-U.S. major media outlets than Yahoo did, while Yahoo used more U.S. major media, U.S. and other news agencies than Google did. In 2008, Google in 2006 used more non-U.S. major media outlets than Yahoo did, while Yahoo used more media than Google did in all other categories. When the search term is ‘Iraq bombing,’ the difference is significant only in 2006 ( $\chi^2 = 97.47$ ,  $df = 5$ ,  $p < .001$ ) and not in 2008 ( $\chi^2 = 7.85$ ,  $df = 5$ ,  $ns$ ). Comparing within each media category, Google in 2006 used more non-U.S. non-major media and U.S. (major and non-major) media outlets than Yahoo did, while Yahoo used more non-U.S. major media, U.S. and other news agencies than Google did.

**Table 5.** Percentages of media of five selected news links on result page by portal

		<i>2006</i>		<i>2008</i>	
		<i>Iraq</i>	<i>Thailand</i>	<i>Iraq</i>	<i>Thailand</i>
Google					
	U.S. agencies	.3	.3	11.0	.7
	Other agencies	.7	11.7	13.3	11.0
	U.S. major media	14.7	.3	13.3	.3
	Other major media	4.0	17.7	10.3	54.3
	U.S. non-major media	68.0	18.7	39.0	19.7
	Other non-major media	12.3	51.3	11.0	14.0
		100%	100%	100%	100%
		(N = 300)	(N = 300)	(N = 300)	(N = 300)
Yahoo					
	U.S. agencies	9.7	9.7	12.0	2.3
	Other agencies	20.0	24.3	19.0	22.3
	U.S. major media	11.0	2.3	15.3	1.0
	Other major media	5.0	12.7	8.3	28.0
	U.S. non-major media	46.7	10.0	37.3	25.3
	Other non-major media	7.7	41.0	8.0	21.0

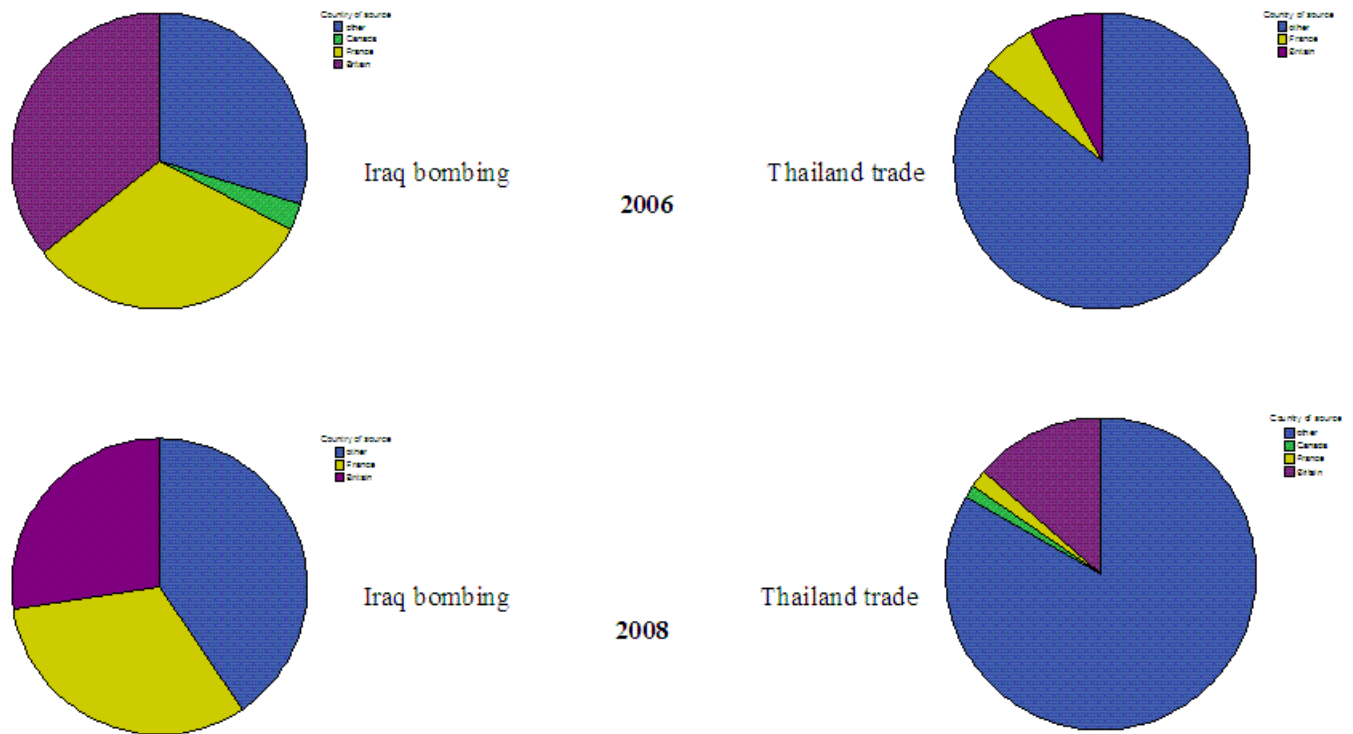
As also shown in Table 5, the result pages of both Google and Yahoo in 2006 used more non-major media outlets (both U.S. and non-U.S.) than major media outlets including news agencies: Google used over 80% of non-major outlets on ‘Iraq bombing’ pages and 70% on ‘Thailand trade’ pages; Yahoo used over 54% and 51%, respectively. In 2008, the percentage of non-major outlets on Google’s pages dropped to 50% for ‘Iraq bombing’ term and about 33% for ‘Thailand trade.’ In the mean time, the decrease in percentage of non-major outlets on Yahoo’s pages was modest: 45% and 46%, respectively. Moreover, in 2006, the largest percentage of outlets on ‘Thailand trade’ result pages belonged to the other non-major media category (Google: 51.3%; Yahoo: 41%). In 2008, non-U.S. major media took over this position (Google: 54.3%; Yahoo: 28%). These figures again clearly show Google and Yahoo’s shift toward more inclusion of major media outlets over the studied period.

In terms of inclusion by percentage, for both news portals, it appears that most of the time the percentage of U.S. media outlets (agencies, major and non-major media alike) on ‘Iraq bombing’ result pages were higher than on ‘Thailand trade’ pages, while the percentage of other media (agencies, major and non-major alike) were higher on the latter pages than the former. As expected, U.S. media accounted for the majority of media outlets on ‘Iraq bombing’ result pages on both news portals in 2006 and 2008, followed by news agencies from other countries. With U.S. media taken out of the picture, around two-third of the media on ‘Iraq bombing’ result pages were from the Great Britain, France, or Canada, while over three-quarters of the media on ‘Thailand trade’ result



pages were from other countries. Figure 5 shows the breakdown of non-U.S. media outlets of Yahoo in 2006 and 2008, which is similar in trend to Google News' pages.

**Figure 5.** Proportion of non-U.S. media outlets by country on Yahoo's result page



Overall, the descriptive data of this study show four major trends: 1) Google News and Yahoo News differed significantly in media inclusion on the front pages, with Yahoo relying on a very limited number of media outlets; 2) the two portals also differed significantly in media inclusion on result pages except for 'Iraq bombing' pages in 2008, with Yahoo relying more on agencies and U.S media while Google on non-major media from the U.S. and other countries; 3) both news portals increased the proportion of major media and decreased that of non-major media between 2006 and 2008; and 4) for both

news portals, the distribution of media outlets was heavily skewed, with very few media outlets used hundreds of times during the studies period while the majority used only a few times (except for the non-major media category on the result pages).

### ***Hypothesis tests***

The first hypothesis suggests a positive relationship between the dominance of the media outlet and position of the news link on the front page – news agencies and U.S. major media are likely to take up higher positions on the front page than other media outlets. Table 6 shows that it is not the case. The test using the combined sample did not yield significance. When examined separately, only with the Yahoo's data in 2006, the correlation test using Spearman's *rho* achieved significance ( $p < .01$ ).

The second hypothesis also suggests a positive relationship between proximity of a news event to the U.S. interest and position of the news link on the front page – events closer to the U.S.' interest are more likely to take up higher positions in the front page than other events. As Table 6 shows, similar results were obtained: The correlation test using Spearman's *rho* only yielded significance ( $p < .05$ ) with Yahoo's data in 2006.

**Table 6.** Spearman's *rho* correlation coefficients for media dominance, event proximity, link position

	<i>Variables</i>	2	3
Combined	1. Media dominance *	-.01	.00
	2. Event proximity **	–	.00
	3. Link position ***	–	–
Google 2006	1. Media dominance *	.01	.00
	2. Event proximity **	–	.04
	3. Link position ***	–	–
Google 2008	1. Media dominance	.00	-.03
	2. Event proximity	–	-.10
	3. Link position	–	–
Yahoo 2006	1. Media dominance	.04	.20 <sup>b</sup>
	2. Event proximity	–	.11 <sup>a</sup>
	3. Link position	–	–
Yahoo 2008	1. Media dominance <sup>X</sup>	–	–
	2. Event proximity	–	-.03
	3. Link position	–	–

\* Coded from 6 (U.S. news agencies) to 1 (non-major media)

\*\* Coded from 5 (U.S. event) to 1 (other single-country event)

\*\*\* Coded from 5 (highest position) to 1 (lowest position)

a.  $p < .05$

b.  $p < .01$

<sup>X</sup> Variable has only one value. Correlation not computed.

Given the time frame of data collection (2006 and 2008), the correlation test using the same sample but excluding news reports about Iraq was conducted to make sure data skewness (i.e., high frequency of news reports about Iraq) did not affect the results. As Table 7 shows, the additional test did not yield different result regarding significance; however, in the case of Yahoo data in 2006, with Iraq taken out of the equation, the correlation coefficient between event proximity and link position became notably higher (.25 compared to .11 of the previous test), implying a stronger positive relationship.



**Table 7.** Spearman's *rho* correlation coefficients for media dominance, event proximity, and link position on front page (Iraq cases excluded)

	<i>Variables</i>	2	3
Combined	1. Media dominance *	-.01	-.02
	2. Event proximity **	–	.03
	3. Link position ***	–	–
Google 2006	1. Media dominance *	.04	-.01
	2. Event proximity **	–	.04
	3. Link position ***	–	–
Google 2008	1. Media dominance	-.00	-.03
	2. Event proximity	–	-.10
	3. Link position	–	–
Yahoo 2006	1. Media dominance	.06	.21 <sup>b</sup>
	2. Event proximity	–	.25 <sup>b</sup>
	3. Link position	–	–
Yahoo 2008	1. Media dominance <sup>X</sup>	–	–
	2. Event proximity	–	-.03
	3. Link position	–	–

\* Coded from 6 (U.S. news agencies) to 1 (non-major media)

\*\* Coded from 5 (U.S. event) to 1 (other single-country event)

\*\*\* Coded from 5 (highest position) to 1 (lowest position)

a.  $p < .05$

b.  $p < .01$

<sup>X</sup> Variable has only one value. Correlation not computed.

The third hypothesis suggests a mediated effect of proximity of a news event to the U.S. interest over the relationship between the dominance of the media outlet and position of the news link such that when the news event is of greater interest to the U.S., media outlets of higher dominance have higher link positions than media outlets of lower dominance, and when the search term is about an event of lesser interest to the U.S. public, media outlets of lower dominance have higher link positions than media outlets of higher dominance. Four ANOVA tests were done separately for Google's result pages in 2006 and 2008, and Yahoo's result pages in 2006 and 2008. As Table 8 shows, with two samples – Google 2006 and Yahoo 2008 – the main effect of media dominance over link position is present but the interaction is not statistically significant in either cases. On the other hand, the main effect of media dominance over link position is not present with the sample of Yahoo 2006, but the interaction is statistically significant,  $F(5,588) = 2.36, p = .039$ . This means that the relationship between media dominance and link position on the result pages of Yahoo in 2006 depends on how close the news event is to the U.S. interest. In other words, the position of media outlets of different categories relative to one another depends on whether search term is 'Iraq bombing' or 'Thailand trade.'

**Table 8.** Two-way analysis of variance of media dominance and event proximity on link position

		<i>Position of news link *</i>		
	<i>Main effects &amp; interaction</i>	<i>F</i>	<i>df</i>	<i>Sig.</i>
Google06	Effect of media dominance	2.63	5	<i>p</i> < .05
	Effect of event proximity	3.70	1	<i>ns</i>
	Interaction: media*event	1.72	5	<i>ns</i>
Google08	Effect of media dominance	1.24	5	<i>ns</i>
	Effect of event proximity	3.93	1	<i>p</i> < .05
	Interaction: media*event	1.37	5	<i>ns</i>
Yahoo06	Effect of media dominance	2.20	5	<i>ns</i>
	Effect of event proximity	.37	1	<i>ns</i>
	Interaction: media*event	2.36	5	<i>p</i> < .05
Yahoo08	Effect of media dominance	3.25	5	<i>p</i> < .01
	Effect of event proximity	5.67	1	<i>p</i> < .05
	Interaction: media*event	1.78	5	<i>ns</i>

\* Coded from 5 (highest) to 1 (lowest).

A post hoc test was employed to further investigate the mediated effect of event proximity on the relationship of each of the six categories of media outlets and position of the news links. The test reveals that only two pairs of interaction exist. As expected, other non-major media outlets were included in higher positions than U.S. non-major media outlets on ‘Thailand trade’ result pages: Regarding link position, the Games-Howell test<sup>v</sup> showed that there was a statistically significant difference between these two



categories ( $p = .018$ ), with U.S. non-major media outlets having a link position mean of 2.37 and other non-major media outlets of 3.15. However, on 'Iraq bombing' result pages, U.S. non-major media outlets were included in higher positions than other major media outlets which reverses the expected order: Regarding link position, the Games-Howell test showed that there was a statistically significant difference between these two categories ( $p = .046$ ), but U.S. non-major media outlets have a link position mean of 3.09 while other major media have a mean of 2.07. Since the post hoc test did not show an expected mediated effect across the categories, the hypothesis was not supported even with Yahoo 2006 data.

The hypothesis tests, using combined data of two portals of two years and separate data of each portal of each year, show none of the three hypotheses were fully supported. International news agencies or big U.S. media might secure top spots on one portal's front pages but not others', and at one time but not others. Similarly, a U.S. or U.S.-related event might dominate the top spots of one portal but not other, and at one time but not others. This goes for a news search as well: Whether the search term is about an event of great interest to the U.S. or about other event of little interest to the U.S. generally does not affect the ranking of media outlets on the result pages.

### ***Discussion***

The empirical findings in this study are probably the first challenge to the propositions put forth by Barzilai-Nahon (2007) in the theory of network gatekeeping regarding the role of the gated relative to the gatekeeper. The fact that empirical data do not support the theory's propositions suggests that perhaps the theory needs to be refined regarding the

gatekeepers themselves. In the case of a web portal, perhaps the most crucial factor to the understanding of the gatekeeping process is how web portals view their market because, after all, they are business entities that work for profit and not purely for the sake of information dissemination. The importance of the gated in light of market profit might be evaluated differently by gatekeepers, hence the difference in treatment of the gated's products. In fact, web portals' front pages and search result pages are shown differently based on the location of the information seekers, and news portals such as Google News have different page versions for different countries.

On another level, these empirical findings reinforce the role of the gatekeepers in the traditional gatekeeping theory. The theory posits that different levels of decision are made in selecting media messages and letting them through their gate (Shoemaker, 1991), resulting in different packages that reach the audience; this study shows this is indeed the case: the all-automated Google gatekeeper and the automated-plus-human-editor Yahoo gatekeeper do let different media messages go through their gate while discarding others. Human editors at Yahoo added an additional gatekeeping level on top of the automated program, resulting in a much narrower news presentation on the pages. The study, however, points to the fact that in the Internet context, it is important to study not only the levels of gatekeeping but also the mechanism by which gatekeeping is imposed on media products. New technologies such as search engine are transcendent – they can be used in different levels by different gatekeepers. How they are used in a way dictates how media messages are treated. The traditional gatekeeping theory does need to evolve to cover this ground in greater depth as the context of gatekeeping has evolved tremendously in the last two decades.

Regarding the issue of bias on the Web, the study's findings challenge previous studies that claimed there was bias in media treatment of web portals. Findings about one web portal should be applied to other portals with caution. A search engine's algorithm is proprietary, and because of the difference in business strategies and different target markets, no two algorithms are completely identical even if two portals employ the same search engine. Therefore, any sweeping generalization about news portals as a single entity might be misleading if not incorrect.

The strength of this study lies in the development of a creative study method that responds to the distinctive nature of news portals. There was no archive of portals' news pages to work on, neither was there a published news index of Google or Yahoo. Using a computer program for data retrieval and summary is a good decision that mitigates the problem. In terms of research validity, the study's large data set makes its overall conclusions about the trends more credible. Furthermore, statistical conclusion is also enhanced and probability of type II error is reduced thanks to large sample size ( $N = 3,600$ ). Even with the test done separately for each portal, the sample size is considerably large ( $N \geq 300$ ).

This study is also an important improvement compared to previous studies on the same subject that usually relied on a single portal or search engine, most often Google, which was an obvious methodological flaw due to lack of what Hargittai (2007) called "inclusive analyses" (p.775) . By employing a comparative and longitudinal approach that pits the all-automated Google against the automated-plus-human-input Yahoo over the years, this study is able to provide a more informative and well-rounded picture of news portals than previous studies. In addition, testing hypotheses using separate data

sets for each portal of each year was a cross-validation step that made the tests more robust than if the tests relied on the combined set or only one of the sets. Overall, the comparative and longitudinal approach greatly enhances the external validity of the study, a strength that should be considered in future research of similar topics.

However, there is room for improvement when it comes to construct development so that future research can achieve higher internal validity. The operationalization of both constructs of interest might need improvement so that they better represent what really goes on in real life regarding media domination and what comprises the U.S. interest. In addition, using specific keywords to represent different values of the construct proximity should be treated with greater caution. It is probably more appropriate to select more neutral keywords rather than using those highly unusual or unique events. Furthermore, the results of the hypothesis test, combined with descriptive data, raise an important point about the necessary distinction between a portal's inclusion and ranking. A news medium might be dominant in total number of occurrences on a portal's pages, but it does not necessarily get the top spots often.

Future research could also look into the issue of news item's origin to see how various news media used news feeds from the news wires, or the issue of access point to see how it affects the inclusion and ranking of worldwide news media.

## References

- Barzilai-Nahon, K. (2008). Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American society for information science and technology*, 59(9), 1493-1512.
- Boyd-Barrett, O. (1997). Global news agencies as agents of globalization. In A. Sreberny-Mohammadi, D. Winseck, J. McKenna, & O. Boyd-Barrett (Eds.). *Media in global context: A reader* (pp.131-144). London: Edward Arnold.
- Boyd-Barrett, O., & Rantanen, T. (1998). *The globalization of news*. London: Sage Publications.
- Couvering, E.V. (2005). Introductory material. *The power of search: Strategy for gatekeeping the Internet*. Unpublished master's thesis, London School of Economics, Britain. Retrieved June 2007, from <http://personal.lse.ac.uk/vancouve/thesis/1-Introduction.pdf>
- Derbyshire, R. (2005). Media and globalization: How has the globalization of the media Industries affected the news media sector? *Radical media online*. Retrieved June 12, 2005, from <http://journalism.cf.ac.uk/2005/MAJS/sjorad1/mediaandglob.html>
- Fortunato, S., Flammini, A., Menczer, F., & Vespignani, A. (2006). Topical interests and the mitigation of search engine bias. *PNAS*, 103(34), 12684-12689.
- Gerhart, S.L. (2004). Do Web search engine suppress controversy? *First Monday*, 9 (1). Retrieved June 2007, from [http://www.firstmonday.org/issues/issue9\\_1/gerhart/](http://www.firstmonday.org/issues/issue9_1/gerhart/)
- Granka, L.A., Joachims, T., & Gay, G. (2004). Eye-tracking analysis of user behavior in WWW search. *Proceedings of SIGIR'04, ACM Press*, 478—479. Retrieved July 10, 2008, from <http://citeseer.ist.psu.edu/article/granka04eyetracking.html>
- Hargittai, E. (2000). Open portals or closed gates? Channeling content on the world wide web. *Poetics*, 27(4). Retrieved June 25, 2005, from <http://www.eszter.com/>
- Hargittai, E. (2003). The digital divide and what to do about it. In D.C. Jones (Ed.). *New economy handbook*. San Diego, CA: Academic Press.
- Hargittai, E. (2007). The social, political, economic, and cultural dimensions of search engines: An introduction. *Journal of computer-mediated communication*, 12, 769-777.

- Hindman, M., Tsioutsoulis, K., & Johnson J.A. (2003). *Googlearchy: How a few heavily linked sites dominate politics on the web*. Paper presented at the annual meeting of the Midwest Political Science Association. Retrieved June 12, 2005, from <http://www.princeton.edu/~mhindman/googlearchy--hindman.pdf>
- Introna, L.D., & Nissenbaum, H. (2000). Shaping the Web: Why the politics of search engines matters. *The information society*, 16, 169-185.
- Kennedy, A. (2005). What clicks with web searchers. Report from the Search Engine Strategies 2005. *Search Engine Watch*. Retrieved June 25, 2005, from <http://searchenginewatch.com/searchday/article.php/3502796>
- Kleinberg, J., & Lawrence, S. (2001). The structure of the Web. *Science*, 294, 1849-1850.
- Lasica, J.D. (2004). *Balancing Act: How news portals serve up political stories*. Retrieved June 25, 2005, from <http://www.ojr.org/ojr/technology/1095977436.php>
- Lee, P. (2005). *Internet news search engines*. Master's thesis, UMI No. 1432200.
- Mowshowitz, A., & Kawaguchi, A. (2002). Bias on the Web. *Communication of the ACM*, 4 (9), 56-60.
- McPhail, T.L. (2002). *Global communication: Theories, stakeholders, and trends*. Boston: Allyn and Bacon.
- Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position, and relevance. *Journal of computer-mediated communication*, 12, 801-823.
- Pew Internet & American Life Project (2005). *Search engine users*. Retrieved March 7, 2006, from [http://www.pewinternet.org/pdfs/PIP\\_Searchengine\\_users.pdf](http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf)
- Schroeder, R., & Kraleman, M. (2005). Journalism Ex Machina – Google News Germany and its news selection processes. *Journalism studies*, 6(2), 245-247.
- Shoemaker (1991). *Gatekeeping*. Newbury Park, CA: Sage Publications.
- Ulken, E. (2005). *A question of balance: Are Google News search results politically biased?* Working paper. Retrieved June 12, 2005 from <http://ulken.com/thesis/googlenews-bias-study.pdf>
- Vaughan, L., & Thelwall, M. (2003). Search engine coverage bias: Evidence and possible causes. *Information processing and management*, 40, 693-707.

Vaughan, L., & Zhang, Y. (2007). Equal representation by search engines? A comparison of websites across countries and domains. *Journal of computer-mediated communication, 12*, 888-909.

Zhao, L. (2004). Jump higher: Analyzing web-site rank in Google. *Information technology and libraries, 23*(3), 108-118.

---

<sup>i</sup> Authority site, a term used in the data analysis field, refers to a site that is linked to by many other sites, or a high traffic site that controls traffic and information flow passing through it.

<sup>ii</sup> The importance of a website in hyperlink is decided by two elements: The quantity of links to web page from other web pages, and quality of the web pages from which the links originated. A site linked to by numerous other sites is considered more important than a site with only two or three links from other sites. A site linked to by prominent sites such as cnn.com is more important than a site linked to by a new or less prominent site.

<sup>iii</sup> The 12-hour interval takes into consideration the fact that the media around the world operate on different time zones. At 9:00am Eastern-time in the U.S., most news of the day is published on news sites. At 9:00pm Eastern-time in the U.S., which is 9:00 am in some Asian countries including the regional media hub of Thailand, news is published online after print time. Since accounting for all time zones is not a viable option, I chose the 12-hour interval as a more practical and realistic option.

<sup>iv</sup> Since the two news portals have different page formats, Google always has a much higher number of media outlets on each page. Comparisons between these two portals were made using total number of distinct (i.e., not counting multiple) media outlets as well as number and percentage of media outlets by frequency of inclusion.

<sup>v</sup> Games-Howell test was used due to violation of the assumption of equal variances showed in the Levene statistic of the interaction test.