

René König and Miriam Rasch (eds), *Society of the Query Reader: Reflections on Web Search*, Amsterdam: Institute of Network Cultures, 2014. ISBN: 978-90-818575-8-1.

# Search Control in China



Min Jiang and  
Vicențiu Dîngă

## Search Control in China



Min Jiang and Vicențiu Dîngă

Min Jiang is an Associate Professor of Communication at UNC Charlotte and an affiliate researcher at the Center for Global Communication Studies, University of Pennsylvania. Min Jiang was a speaker at the second edition of the Society of the Query conference in Amsterdam, 7-8 November 2013. She participated in the session called 'Search Across the Border', where she talked about borders on web search in China, expanding on her previous research on search results in China.

**Vicențiu Dîngă: Firstly, I'd like to talk about your work involving the comparative study of Baidu and Google's search results in China. Could you tell me more about this research?**

Min Jiang: The impetus for this research came from my own varied experiences with search engines in China, the U.S. and elsewhere. How is search experienced differently became a question that I wanted to explore. Search engines are interesting media to me. They are an important interface between the user and available knowledge and information. More interestingly, they are also popularly regarded as neutral and almost god-like while search's political and commercial nature is often downplayed. I explored this aspect in the Chinese context. In an article published in *New Media & Society* I compare Baidu's and Google's search results in China, focusing on a few aspects: search filtering, ranking, overlap, and bias.<sup>1</sup>

First, we know that filtering has always been an issue with the Chinese internet, and search filtering is a part of it. Google, upon entering China in 2006, consented to filtering. But in January 2010, Google announced that it would stop censoring its search results in mainland China because of security threats. Subsequently Google moved its servers to Hong Kong because the region is a free speech zone in China. Many suspected the censorship burden shifted from Google to the 'Great Firewall', a filtering infrastructure erected at the Chinese borders so that international information coming through can be filtered ('The Great Firewall of China' is the colloquial term for the 'Golden Shield Project'), but nobody knew how effective the Great Firewall was. Part of my data shows that there were quite complex filtering patterns in China. Top 10 search results from both Google China and Baidu were collected inside mainland China in late 2010 using 316 popular event keywords. The table below indicates that after Google

---

1. Min Jiang, 'The Business and Politics of Search Engines: A Comparative Study of Baidu's and Google's Search Results of Internet Events in China', *New Media & Society*, 22 April 2012, <http://nms.sagepub.com/content/early/2013/04/18/1461444813481196.full>.

left mainland China, the filtering burden indeed shifted to the Chinese state. While inaccessibility of Baidu's search results was caused mainly by broken links, for Google, it was largely because of filtering by the Great Firewall. Baidu also performs this task of filtering as Chinese laws mandate content filtering by commercial operators within its borders. So, Baidu filters too, but its filtering patterns are much subtler and more nuanced than that performed by the Great Firewall.

	<b>Baidu</b>	<b>Google</b>
Total number of links	3160	3160
Links inaccessible from China	171	400
Bad links	125	91
Links inaccessible due to search engine filtering	22 (partial filtering or blockpage by Baidu)	10 (by Google SafeSearch auto-block)
Links inaccessible from China but accessible from the U.S. (blocked by GFW)	24	299 (among them, 180 'connection reset')
Links inaccessible due to GFW blocking	24	299

Table 1. Distribution of Baidu and Google's Link Accessibility (2011).

Second, in terms of search results overlapping and ranking, I found Google and Baidu are actually quite different. Based on the search results of 316 popular Chinese events, the overlap, or the appearances of the same URL in top 10 search results from both search engines, is quite low. Maybe this is not so shocking to those of us who study search engines, but for the general public, an overlap of search results between Google and Baidu of less than 7% might come as a big surprise. This tells us that based on the search engine you use, your search results and the knowledge you obtain as a result can vary in significant ways. When it comes to news and current events, search engines can be quite important in shaping what we know.

With regard to search bias, I focused on self-bias, or the extent to which a search engine links search results to its own content. This is because, as we know, Google and Baidu are no longer only search engines; they offer many other related information products and services. For instance, Baidu owns Q&A websites and real time messaging services. It has incentives to direct searchers to its own content, with effective regulation or strong competition absent. At the time of my study in late 2010, it turned out that Baidu rarely links to Chinese Wikipedia or other competitors in China, but often to Baidu Baike, its own Wikipedia-like service. On the other hand, Google back then was more likely to link to Chinese Wikipedia. It seems at times that Chinese Wikipedia is Google's default choice even though Chinese domestic competitor Hudong Baike was more popular with many more entries and users at the time. The latter case is not about Google's favored treatment of its own service, but its preferences certainly shaped users' information experiences and outcomes. In a follow-up study with a component examining Google's linking patterns to Wikipedia-type services, I found Google has

since shifted its strategy.<sup>2</sup> Based on the longitudinal data I collected in 2011 and 2012, Google no longer seems to privilege Chinese Wikipedia and instead has given Baidu Baike and Hudong Baike an almost equal chance in mainland China.

Search bias is a real concern. Even Chinese official media CCTV has exposed Baidu for punishing advertisers who refused to pay for their search rankings. I think this issue will persist as long as search engines remain a critical information access point in our societies. More and more people will start to realize this is a huge problem. While China's political and market situation may be somewhat unique, search bias and unfair business practices can occur in every information society and market.

**VD: How do these search engines behave throughout China?**

MJ: Although search results can vary for a variety of reasons, regional differences are notable. Search results in Hong Kong, Taiwan, and Macau, for instance, differ in important ways from those retrieved from mainland China. This is not only because of politics, but also because of linguistic and cultural differences among other things. As an example, Hong Kong, previously a British colony, was returned to mainland China in 1997 but the two regions remain distinct under the 'one country, two systems' principle (where Hong Kong practices a capitalist economic and political system and mainland uses a socialist one). In Hong Kong, local laws continue to protect free speech – that's why Google moved its servers there and to this day is able to remain in 'China'. The legal systems in these regions differ in marked ways from mainland China's. Linguistically, Hong Kong natives continue to speak Cantonese and Taiwan natives Taiwanese. Traditional Chinese characters are more commonly used in Hong Kong, Macau and Taiwan rather than simplified Chinese characters, the latter used more dominantly in mainland China. In all its complexity, the notion of China may include Hong Kong, Macau, and often Taiwan (although mainland and Taiwan still have unresolved territorial and political disputes). These historical legacies continue to shape how search is used and regulated today. In the Chinese case, search tends to conform to these linguistic, cultural and legal boundaries rather than transcending them. Search is re-territorialized rather than de-territorialized. My colleague Han-Teng Liao from Oxford's Internet Research Institute has a paper that discusses the different patterns of search results that one gets in Chinese-speaking regions including mainland China, Hong Kong and Taiwan.<sup>3</sup>

**VD: In your presentation for the Society of the Query conference you stated that 'the popular depiction of the search engine as a borderless, global medium is an illusion'. Can you expand on that?**

MJ: Yes. I believe there is a huge gap between what we think search can offer versus how we use it in reality. We have this dream, almost a fantasy, that search engines are

- 
2. Min Jiang, 'Search Concentration, Bias, and Parochialism: A Comparative Study of Baidu, Jike, and Google's Search Results from China', presented at the 41st Research Conference on Communication, Information and Internet Policy (TPRC), 27-29 September 2013, George Mason University School of Law, Arlington, VA, USA.
  3. Han-Teng Liao, "'Wikipedias' (or its Copycat) Dominate 'Chinese' Search Engine Result Pages (SERPs)", 6 May 2013, <http://people.oii.ox.ac.uk/hanteng/2013/05/06/wikipedias-or-its-copycat-dominate-chinese-search-engine-result-pages-serps/>.

a borderless and global medium. In reality, I think our experiences are very much conditioned by our location, by who we are, how we search, and what we know. In many ways, search is more local or parochial, if you will, rather than cosmopolitan or global. Search engines are popularly mentioned in the same breath as the internet as part of a *global village*, a McLuhanian dream built on an information system radiating throughout the world, producing positive outcomes. As with any new technology, however, as time went by, more and more governments also started to figure out how to use law and other means to set the boundaries for search and in extreme cases to filter any undesirable content. Structurally, we have a global digital infrastructure that is more or less connected, but in terms of content and information flows, I think it's quite a local or national experience for many, especially in countries such as China. At the inception of the internet and search engines, those who were able to use them were more educated and cosmopolitan to begin with. Today, in China alone, there are almost 600 million internet users and 450 million searchers. For many, their search experiences are not at all as borderless as projected.

Borders are frames. People think about them often in terms of physical borders and territory. On the internet, there are no physical borders; borders are invisible. In the realm of search, various search engines employ complex criteria to collect, sort, rank, and present results to us, adopting certain factors and excluding others, operating largely in secrecy and expanding their collection of user data to ever more spheres. Eli Pariser describes such criteria or algorithms as 'filters' and the resulting state of individualized information as 'filter bubbles'.<sup>4</sup>

Geo-linguistic borders, the kinds of existing differences in location and language, for instance those between Chinese mainland, Hong Kong, and Taiwan as noted previously, certainly form the basis for distinct sources and domains of information and knowledge from which search engines may crawl. They are reflected in search results. Politics, regional or national, could also play an important role, especially in places where freedom of political speech is not guaranteed. Another aspect of 'borders' has to do with technology itself in the form of geo-location. If you are located in Amsterdam, your search results are more likely to receive localized results based on your IP address. This aspect has become more and more prevalent over time, as search engines do not offer universal results to a query anymore. The degree and scope of personalization, however, remains largely a mystery. It has been revealed that Google uses over 70 different 'signals' to personalize results. Broadly, Google defines a person in three aspects: your geographical location, search history and your relationships or social networks. These are frames or 'borders' imposed on the user. A last aspect I want to emphasize is economic incentives for localized delivery of information. A lot of our search behavior is 'local' – whether you want to get a haircut or buy a product at a nearby grocery store. Localized advertising is pushing for localization in search results. While localization makes sense in some cases, it can turn out badly in others where the need to transcend parochialism is more important. In general, a variety of factors aside from personal idiosyncrasies – linguistic/cultural, political, technological and economic – have come to shape our search results to be more local or parochial.

---

4. Eli Pariser, *The Filter Bubble: What the Internet Is Hiding from You*, New York: Penguin, 2011.

**VD: You mentioned that possible alternatives could be offered to reimagine what search engines could become. What possible alternatives do you have in mind?**

MJ: When it comes to news, to current events, it is quite inadequate for the user to only receive results that are based on location. China is a very extreme example. In some cases, you do not get any alternative perspectives at all on critical events. During the Arab Spring, search results were filtered in such a way that the only news you got in China were from state sanctioned sources. It tended to focus on the government's efforts of rescuing the Chinese from the troubled regions and China's investments in those areas. Search results made little to no mention of the protests or demonstrations for fear of domestic repercussions. When these did get mentioned, they were presented as chaotic and threatening. Stability trumps everything. That is all you get. Given the current political arrangement in China, change will be hard to come by. Search reflects rather than challenges the political make-up of the regime.

But people elsewhere may have more options. Imagine, if Google can personalize your search results, then why doesn't Google give you the option to personalize your own results? For instance, why can't I design my own filters if I choose to be more internationally minded? Search engines could give users more freedom in the design of their own search results and the kinds of information they want to receive. Google, for instance, has been experimenting with personalized Google News since 2011 where 'borders' could be rendered less relevant by users themselves. Certainly, by choosing one's own filters, one has to live with a set of self-imposed 'filter bubbles'. But this may allow for a greater degree of self-determination rather than an opaque set of algorithmic rules over which we have no control. There have been suggestions that users should minimize or completely abandon the use of search engines because using them would subject users to a set of opaque algorithms (not unlike a robotic monster) that feeds on, lives off, and profits from our incessant use. This is a difficult dilemma. Will people stop using search engines entirely? Probably not. But could there be more algorithmic scrutiny, user control and privacy protection? Much more likely. In fact, there are alternative search engines out there (e.g. DuckDuckGo and Blekko) that offer users more privacy protection and options.

There is also a lot of room for independent evaluation of search and search quality. Many of us use one search engine on a regular basis, be it Google in much of the world or Baidu in China. As the amount of information explodes, the job of delivering better information becomes harder, not easier. The notable differences between search results delivered to the same person from different search engines naturally beg such questions as 'which is better?' and 'based on what standards?' The problem, however, is that perfectly ranked search results for any given query do not exist. Algorithms are a kind of simulacra in the Baudrillardian sense. Yet, we can still compare search engines and reveal their preferences or prejudices, and in some cases even abuses of their surveillance and monopolist power.

**VD: What about the Chinese state search engines, how do they perform?**

MJ: In China, there are three state-run search engines. One of them is well known as the first national Chinese search engine. It's called Jike, which in Chinese means 'instant' or 'immediate'. Some call it 'geek'. Its parent company is *People's Daily Online*, which is publicly listed on Shanghai Stock Exchange. Another one is Panguso; it's a search engine backed by Xinhua News Agency (a state news agency)

and China Mobile (also a state-owned enterprise publicly listed on both the New York Stock Exchange and the Hong Kong Stock Exchange). The third one is a search engine associated with CCTV, China's official television station, now with an online presence through CNTV. Recently, Jike got into financial troubles and has merged with Panguso.

So diverse parties of interest are at play here and the reasoning behind state-sponsored search engines is quite complex and convoluted. I explored the phenomenon of 'national search engines' in a paper that will appear in *Policy & Internet*. First, security was a major concern. All three state-backed search engines were put into play immediately after Google left mainland China as a response to Secretary Hilary Clinton's 'Internet freedom' speech. It was clear from publicly available Chinese Communist Party's documents that China's top leaders consider 'internet sovereignty' and control over information within Chinese borders a top priority. They attached great importance to the 'management' of the internet in China and understood the centrality of search engines in an information society.

Second, such an acute sense of security was heightened by two subsequent pivotal events – the Arab Spring and the Snowden revelations, both of which have prompted a global trend of re-nationalization of the internet. Repercussions of the Arab Spring are still unraveling but the quick toppling of several dictatorships sent a palpable jolt to Chinese leaders in Beijing who were obsessed with 'stability'. National search engines were part and parcel of increased government control asserted over the Chinese internet. In the post-Snowden world, there were several important signs of re-nationalization of the internet, including Russia's development of its own national search engine *Sputnik*, and discussions of a 'national internet' in both Germany and Brazil. To what extent 'national internets' will materialize in Germany and Brazil is not clear but in China, there is already a *de facto* 'national internet'.

Third, financial considerations also played an important role in China's state search engine experiment, where the Party wishes to profit from China's surging information and search markets. Since China started to implement economic reforms in the late 1970s, state media have become more independent financially, although not quite free from the state's ideological management. Chinese mainstream media, a euphemism for state-owned media, absolutely dominate China's traditional media from books and magazines to radio and TV, however not so much in online media. State propaganda has been trying to figure out how to engage users and do propaganda work online. *People's Daily*, the party's mouthpiece, has developed a successful website *People's Daily Online* whose online traffic is among the top 20 in China (based on Alexa data). It seems that, with national search engine Jike, *People's Daily Online* hopes to expand its sphere of influence and become more profitable financially. However, I think authorities may have underestimated the effort it takes to develop a commercially viable search engine. Jike failed technologically, strategically and managerially. Jike's failure, however, has not put an end to state search engine experiments in China. After Jike's merger with Panguso, the future of state search engines remains highly uncertain but is undoubtedly subject to both politics and competition.

**VD:** You mentioned the Great Firewall of China, and I remember reading somewhere that it is suspected that regimes such as Cuba, Zimbabwe, or Belarus have

**obtained surveillance technology from China. So it seems that China actually exports this kind of approach.**

MJ: That's right. When we talk about Chinese search engines or the Chinese internet in general, we tend to neglect or underestimate China's reach overseas. Apart from domestic growth driven by nearly 600 million Chinese internet users, China's rise technologically, politically, economically, and military is also manifested overseas. In fact, Chinese technology companies, state-owned or commercially operated, seek to grow abroad through international trade as many others do. This includes a unique category of surveillance and filtering technologies that the state or state-related companies have developed and wanted to export. While China has previously imported surveillance technologies from such Western companies as Cisco, Microsoft, and Nokia, it has been developing its own surveillance technologies that might end up in other countries as you mentioned. Chinese companies such as Huawei (a telecommunication company and the largest producer of telecommunications equipment in the world) have been quite successful in exporting tech products. Lately, news reports surfaced that Huawei plans to sell technologies to Iran to help 'clean' its internet. While many of Huawei's low-cost, functional tech products can aid growth and development in some of the world's poorest areas in Africa, South America, and Asia, it is also competing against other Western companies in selling filtering technologies to authoritarian regimes.

**VD: What do we know about the way that users deal with these censorship regimes? Is there research being done on how they try to counteract this?**

MJ: This issue has attracted the attention of a lot of Western countries and reporters. Sure, some Chinese users are aware that their search results are being filtered. Users who are knowledgeable and tech-savvy are able to use proxy servers and VPNs to route around censorship. But we do not know how many people are aware of this issue or how many actually care about their search results enough to seek alternative search engines and processes. I believe the majority of users either don't know or don't care. Maybe their information needs are met by the filtered results. Maybe they will not bother paying extra money or doing extra work just to get 'better' search results. I am not too optimistic about it. People who have a lot of business or academic connections with the outside world are more likely to employ alternative strategies. In fact, most foreign businesses in China have been using proxies or VPNs for a long time. For them, it's a standard operation. Also, individuals who rely on a daily dosage of alternative information from valuable external sources such as *The New York Times*, the Chinese version of *Wall Street Journal*, BBC Chinese or Twitter are likely to use circumvention tools on a more regular basis.

**VD: We talked about Google and Baidu, the biggest players in China. I recently read that Bing was reportedly censoring results from searches conducted in simplified Chinese<sup>5</sup>, so the ones used in mainland China, but not the ones in traditional Chinese characters used in Hong Kong, Taiwan, or Macau.**

MJ: I believe that confirms what we talked about earlier. Bing's censoring of its search results in mainland China is to comply with (mainland) Chinese laws. It does not censor results in traditional Chinese perhaps because most of the people who use traditional

---

5. Pete Cashmore, 'Bing Accused of Censoring Simplified Chinese Language Searches', Mashable.com, 21 November 2009, <http://mashable.com/2009/11/21/bing-chinese-queries/>.



Chinese are from Hong Kong, Taiwan, or Macau, regions with different speech regulations. But it would be interesting to learn, for instance, if mainland searchers using traditional Chinese for their queries might get uncensored results. In many ways, censorship is a language game.

**VD: One more thing: what is your opinion on the importance of search engines? Can Google and other search engines substitute knowledge somehow? Or are they more like a tool that shapes and distributes knowledge through their rankings and mechanisms? And if these search engine mechanisms can be controlled by political authorities and interest groups, how does that affect knowledge?**

MJ: Many people equate search engines with knowledge, and probably most search engines want people to think that way. A lot of us use search engines to access information and search engines do matter. Wikipedia is so popular precisely because search engines like Google have made it a default 'knowledge site.' By doing so, search engines like Google define to some extent what constitutes useful information or knowledge. Every search engine wants to crawl and index as much online content as possible, but it's impossible to index everything. In 2008 Google announced it had indexed one trillion unique URLs. Lately the number has gone up to 60 trillion. But even that is a small fraction of what's available out there and it is not clear that all that gets indexed can be considered useful knowledge.

It is misleading, I think, to equate search engines with knowledge. To me, knowledge means a set of facts, truths and beliefs learned throughout time. Knowledge can be gained through experience and human interactions. People have always learned through curiosity, trial and error, intuition, rational thinking, and we learn a lot from others. Since learning is situated in various social and institutional contexts, knowledge can be considered the result of a highly social and culture-specific process. Lately, we have been able to locate information much faster through a process facilitated by search engines, but it is dangerous to think knowledge is a mere result of filtering and algorithms.

The equation of search with knowledge, I suspect, may have something to do with our fear in dealing with information overload and our huge investments in information technologies. By equating search with knowledge, we take comfort in the fantasy that the world's information and knowledge is at our fingertips. The linear progression 'data-information-knowledge-wisdom', to me, represents our wishful thinking rather than a real process of acquiring knowledge that is always messier, harder and more serendipitous. This is not to deny the usefulness of search, but to point out knowledge is much more than search filtering and algorithms.

Search engines, political authorities, business interests and powerful interest groups can certainly impact the process and mechanisms of search. In the end, whatever results Google, Baidu, or any search engine presents to the user depends very much on what the search engine can crawl and index, what it desires or is allowed to present. It's a highly problematic process. Unfortunately, we don't have a very good understanding of its specifics. In certain regimes, the information sought by users is sometimes hidden and forbidden. Steep barriers still exist, preventing people from participating fully in global conversations and knowledge productions. Much needs to be done to open dialogues and increase access to information and knowledge without losing sight of the need to preserve user privacy and integrity.

## References

- Cashmore, Pete. 'Bing Accused of Censoring Simplified Chinese Language Searches', Mashable.com, 21 November 2009, <http://mashable.com/2009/11/21/bing-chinese-queries/>.
- Jiang, Min. 'The Business and Politics of Search Engines: A Comparative Study of Baidu's and Google's Search Results of Internet Events in China', *New Media & Society*, 22 April 2012, <http://nms.sagepub.com/content/early/2013/04/18/1461444813481196.full>.
- Jiang, Min. 'Search Concentration, Bias, and Parochialism: A Comparative Study of Baidu, Jike, and Google's Search Results from China', Presented at the 41st Research Conference on Communication, Information and Internet Policy (TPRC), 27-29 September 2013, George Mason University School of Law, Arlington, VA, U.S.
- Liao, Han-Teng. '"Wikipedias" (or its Copycat) Dominate "Chinese" Search Engine Result Pages (SERPs)', 6 May 2013, <http://people.oii.ox.ac.uk/hanteng/2013/05/06/wikipedias-or-its-copycat-dominate-chinese-search-engine-result-pages-serps/>.
- Pariser, Eli. *The Filter Bubble: What the Internet Is Hiding from You*, New York: Penguin, 2011.