

According to Local Laws and Corporate Policies:  
Internet Filtering on Google.cn 2008

by

Katie Puxley

Submitted in partial fulfilment of the requirements  
for the degree of Master of Library and Information Studies

at

Dalhousie University  
Halifax, Nova Scotia  
December 2015

© Copyright by Katie Puxley, 2015

*For Dave.*

*Amo te.*

# Table of Contents

List of Tables .....	v
List of Figures.....	vii
Abstract.....	viii
List of Abbreviations Used.....	ix
Glossary .....	x
Acknowledgements.....	xii
Chapter 1 Introduction.....	1
1.1 Research Statement .....	3
1.1.1 Research Questions.....	3
1.1.2 Research Context .....	4
1.2 Background .....	6
1.2.1 The Internet in China .....	6
1.2.2 Sensitive Matters: Falun Gong and Tibet .....	10
1.2.3 Google in China .....	12
Chapter 2 Literature Review.....	13
2.1 Removing Sensitive Materials from the Internet .....	18
2.2 Internet Filtering in China.....	21
2.3 Circumventing the Censors .....	26
2.4 Researching Internet Filtering in China .....	27
2.4.1 The Google.cn Years 2006-2010 .....	30
2.4.2 Overblocking.....	34
2.5 Conclusion.....	35
Chapter 3 Methodology.....	36
3.1 Keyword Selection.....	36
3.1.1 Variables .....	37

3.2	Data Collection Instrument .....	37
3.3	Procedure.....	43
3.3.1	Pilot Study.....	43
3.3.2	Volunteer research assistants .....	43
3.3.3	Data Collection .....	44
3.4	Data Analysis .....	46
Chapter 4	Results.....	48
4.1	Frequency of Filtering of Search Results.....	48
4.2	Consistency of Filtering .....	50
4.3	Frequency of Filtering by Language .....	53
4.4	What was Filtered?.....	55
4.4.1	Summary of Unique URLs Filtered in Google.cn .....	56
4.4.2	Overblocking.....	60
Chapter 5	Discussion.....	63
5.1	Research Questions .....	64
5.1.1	Pervasiveness of Filtering.....	64
5.1.2	Consistency of Filtering.....	66
5.1.3	Language and Filtering .....	67
5.1.4	Content Removed and Overblocked.....	68
5.2	Research Design Issues .....	70
5.3	Contributions.....	71
5.4	Future Research.....	73
Chapter 6	Conclusion .....	74
Appendix A	Keywords .....	77
Appendix B	Data Collection Screen Shots.....	83
Appendix C	Results Tables .....	86
References	.....	95

## List of Tables

Table 1 Keyword counts by topic and language .....	37
Table 2 Variable descriptions .....	37
Table 3 Cochran Q mean values for Chinese keywords search by run .....	51
Table 4 Cochran Q mean values for English keywords search by run .....	52
Table 5 McNemar mean values for English (EN) and Chinese (ZH) keyword searches by run .....	54
Table 6 Total number of unique URLs by organization and domain from all filtered keyword searches .....	57
Table 7 Number of unique URLs by organization and domain from filtered Falun Gong keyword searches .....	59
Table 8 Number unique URLs by organization and domain from filtered Tibet keyword searches .....	59
Table 9 Organisations with anti-Falun Gong/Tibet material affected by overblocking ...	60
Table 10 Organisations unrelated materials affected by overblocking.....	61
Table 11 Number and percentages of unique URLs that were simultaneously filtered and available in site search results.....	62
Table 12 List of keywords with translation and topic.....	77
Table 13 Count of always filtered Falun Gong URL site search results grouped by organisation.....	86
Table 14 Count of sometimes filtered Falun Gong URL site search results grouped by organisation.....	88
Table 15 Count of always filtered Tibet URL site search results grouped by organisation .....	89

Table 16 Count of sometimes filtered Tibet URL site search results grouped by organisation.....	90
Table 17 Count of overblocked anti-Falun Gong/Tibet URLs grouped by organisation and domain.....	91
Table 18 Count of overblocked unrelated URLs grouped by organisation .....	92
Table 19 Overblocked URLs that were simultaneously filtered and available in site search results .....	93
Table 20 Sensitive URLs that were simultaneously filtered and available in site search results .....	94

## List of Figures

Figure 1 Google.cn screen capture: Google.cn filtering statement.....	2
Figure 2 Data collection web tool: Section 1 login screen .....	38
Figure 3 Data collection web tool: Section 2 Google.ca keyword search .....	40
Figure 4 Data collection web tool: Section 3 Google.cn keyword search .....	41
Figure 5 Data collection web tool: Section 4 Google.cn URL site search(s) .....	42
Figure 6 Percentage of all keyword searches filtered at least once .....	49
Figure 7 Frequency of keyword search filtering by language and topic.....	50
Figure 8 Cochran Q mean values for Chinese keywords search by run .....	52
Figure 9 Cochran Q mean values for English keywords search by run.....	53
Figure 10 Proportion of filtered English (EN) and Chinese (ZH) keywords per run .....	55
Figure 11 Percentage of filtered unique URL site searches.....	58
Figure 12 Comparison of Google.cn keyword search filtering rates to proportion of top ten Google search results inaccessible in previous studies.....	66
Figure 13 Google.cn screen capture: Filtered site search, results removed.....	83
Figure 14 Google.cn screen capture: Filtered site search, no sites indexed .....	83
Figure 15 Google.cn screen capture: Unfiltered site search, result returned .....	84
Figure 16 Google.cn screen capture: Unfiltered Google.cn site search, no sites indexed	85

## **Abstract**

From 2006 to 2010 Google Inc. operated a filtered regional search engine specific to China: Google.cn. Google's regional database complied with the People's Republic of China's laws on sensitive and secret information and was launched so that Google Inc. would be allowed to operate locally.

The objective of this research is to gain a better understanding of the pervasiveness and consistency of the Internet filtering practices that Google employed in Google.cn. This study a) tracked the presence or absence of Google's filtering statement within Google.cn search results, and b) compared the top ten results returned from Google.ca with those of Google.cn to determine whether absent URLs were available in the Google.cn database or were filtered.

A list of 180 English and Chinese keywords was compiled, all pertaining to Falun Gong and Tibet, which are known to be sensitive topics in China. This set of keywords was searched seven times over the course of a month in May 2008. Initial findings revealed that filtering was pervasive on Google.cn: 81% of keyword searches were subject to filtering. The filtering was consistent in each run of keyword searches. Of the compiled list of URLs 38% were subject to filtering; 6% of the 38% were the results of overblocking.

While Google Inc. was transparent in notifying users that filtering had taken place, the results of this study demonstrate that a significant amount of content was subject to filtering and that it is extremely difficult to examine what content was filtered.



## List of Abbreviations Used

CCP: Chinese Communist Party

*EAIinC: Empirical analysis of Internet Filtering in China* (Zittrain & Elderman, 2003)

GFW: Great Firewall of China

*IFinC: Internet filtering in China 2004-2005: A country study* (OpenNet Initiative, 2006)

IT: Information Technology

ONI: OpenNet Initiative

PRC: People's Republic of China

RA: Research Assistant

URL: Uniform Resource Locator

## **Glossary**

Google.ca: Google's Canadian regional search engine.

Google.com: The worldwide homepage of Google Search and the domain of original Google search engine for the United States of America.

Google.cn: Google's Chinese regional search engine, which from January 2006 until January 2010 was filtered in accordance with Chinese laws on sensitive and secret information.

Google Inc. (Google): Google Inc. is a multinational technology company based in the United States of America. Their first and best known product is their Internet search engine, Google Search. Since its incorporation in 1997, Google has expanded its products and services to include many other technologies and Internet applications, such as cloud computing, software, and hardware. (Google, n.d.)

Great Firewall of China (GFW): "A barrier to keep its Internet users from dealing easily with the outside world ... part of a larger, complex structure of monitoring and censorship [known as] the 'Golden Shield Project'." (Fallows, 2008, p. 66)

Internet Service Provider (ISP): The party providing connectivity to the Internet. (Berners-Lee, 1999, para. 35)

Keyword: any informative word that describes or is part of the content of the records contained in a search engine, i.e., in the title, content, or URL, etc.

OpenNet Initiative (ONI): "OpenNet Initiative is a collaborative partnership of three institutions: the Citizen Lab at the Munk School of Global Affairs, University of Toronto; the Berkman Center for Internet & Society at Harvard University; and the SecDev Group (Ottawa) [whose] aim is to investigate, expose and analyze Internet filtering and

surveillance practices in a credible and non-partisan fashion” (OpenNet Initiative, n.d.-b, para. 1-2)

Overblocking: “[Overblocking], where sites with superficial similarities to those with sensitive material, but different content, were filtered” (OpenNet Initiative, 2006, p. 23).

Pervasiveness of filtering: measure of the extent to which the Google.cn filtering statement appeared in keyword or URL site searches

Query: a request to a search engine by a user for records that relate to a specific keyword

Run: single, complete execution of a computer program; in this case, searching the full set of 180 keywords in Google.cn and Google.ca as laid out in Section 3.3.3 Data Collection.

Search engine: some combination of hardware and software with a stored index of webpages that is capable of returning lists of pages that match keyword queries.

Uniform Resource Locator (URL): “The string (often starting with http:) that is used to identify anything on the Web” that maybe subject to change (Berners-Lee, 1999, para. 83).

## Acknowledgements

Thank you to the many people whose support and assistance contributed to the production of this thesis. First and foremost, my thanks goes to Dr. Keith Lawson, my thesis advisor, without his dedication and support throughout the years, this project would not have come to fruition. Many thanks to my committee members, Dr. Jamie Blustein, who has been committed to this project since its inception, and, Dr. Jennifer Grek Martin, for coming on to see it through to completion. Special thanks to Dr. Lori McCay-Peet, for her expertise, support and guidance, which greatly improved this thesis.

I am greatly indebted to the many people that gave their time to help me with technical support, data collection, and research advice over the many stages of this thesis. I would like to thank the volunteer research assistants, who assisted with data collection: Alison Eaton, Amy Goldlist, Andrea Norman, Bradford MacSwain, Christina Macdonald, Dave Puxley, David Neatby, Deborah Hicks, Ed Leach, Gaie Orton, Gavin Anderegg, Jennifer Bawden, Juanita Smith, Karl Siewert, Kate Snow, Keith Lawson, Lesley Penney, Lesley Roberts, Liam O'Brien, Megan Fitzgibbons, Melanie Frappier, Polly Ross, R Bartlett, Robin C, Sachi Gibson, Samantha Copeland, Suzanne Sheffield, Tasya Tymczyszyn. Thank you to my translators, Jin Wie (Will) and Liu Wan-Ni (Stephanie), for their expertise. A very special thank you goes to my roommates at Harvard St. (in particular Lesley Penney) who not only did the majority of the data collection, but were so efficient at searching that Google blocked them on suspicion of being a bot. A huge thank you goes to Gavin Anderegg and Shamus Peveril for all their work on the programming and for answering the unending stream of questions I've asked ever since.

I also need to thank the many people that helped behind the scenes so that I could continue my work, there are simply too many to name everyone here but I would particularly like to recognise: Joan Carling and Jennifer Bawden for being my family support when my immediate family was far away. Many thanks to the Mount Saint Vincent Library staff for all the support, mentoring, advice and comments. The Mount Library is one of the most congenial and supportive places at which I've ever had the opportunity to work; thank you all.

Finally, I want to gratefully acknowledge the unfailing support of my family. I am extremely blessed to have such wonderful support. Thank you to my mom, Gaie, and my in-laws, Susan & Alan, for your unflagging confidence in me. My husband Dave is the ultimate reason this project was completed. For his endless motivation (even when I didn't want it), for doing considerably more than his fair share on the home front, and for his willingness to always read one more draft, I am truly grateful. Lastly I would like to thank my son, Simon, for his patience when I had to get my homework done and his encouragement to "get your thesis done so it will be finished".

## Chapter 1 Introduction

Google.cn presents to users a clear notification whenever links have been removed from our search results in response to local laws and regulations in China. We view this as a step toward greater transparency that no other company has done before. (Schrage, 2006, The Decision: What Google Is Doing in China section, para. 3)

From 2006 to 2010 Google Inc. operated a regional search engine specifically for China: Google.cn. During this period the regional database complied with the People's Republic of China's (PRC) laws on sensitive and secret information so that Google Inc. would be allowed to do business locally in China. Google's decision to create a filtered search engine and to work within the Chinese censorship rules gave fresh life to the debate about online censorship.

At the time it was launched, Google announced that it would filter Google.cn. The pervasiveness of that self-censorship was, however, the subject of much speculation. Google was open about the fact that Google.cn was filtered from 2006 to 2010; indeed, filtered results had a statement at the bottom of the page notifying users that results had been removed from the search results (see Figure 1).

The focus of this study is not China's policies or practices but rather Google's self-censoring activities in compliance with the laws of the PRC. Using keywords that are known or reported to have been censored by the PRC to search both Google.cn and Google.ca (Google's Canadian regional search engine), this study documents the pervasiveness and consistency of the filtering on Google.cn. All the keywords used in the study relate to Tibet and Falun Gong, topics chosen specifically because they were known

to be considered highly sensitive by the PRC, and therefore were likely to be filtered by Google.cn: “[in] China, Web sites that represent the Falun Gong and the Tibetan exile groups are widely blocked” (Faris & Villeneuve, 2008,c p. 12). Moreover, both Chinese and English language keywords relating to each topic were included to assess whether language had any effect on the filtering.

Figure 1 Google.cn screen capture: Google.cn filtering statement



Previous studies conducted by the OpenNet Initiative (ONI) used various indicators to track Internet filtering by nation states, notably *Empirical analysis of Internet filtering in China (EAIFinC)* (Zittrain & Elderman, 2003) and *Internet filtering in China 2004-2005: A country study (IFinC)* (OpenNet Initiative, 2006) focused on China’s Internet filtering practices. The present study is adapted from ONI methodology, primarily from the *IFinC* study, in which Google.com results from outside of China were compared to Google.com results from inside of China to determine the extent of China’s Internet censorship.

## **1.1 Research Statement**

The objective of this research is to gain a better understanding of Google's Internet filtering practices with respect to Google.cn from an end user perspective by tracking the information displayed in Google.cn search results. This study examines Google.cn's filtering practices on the topics of Falun Gong and Tibet in English and Chinese by a) tracking the presence and absence of Google's filtering statement within Google.cn keyword search results, and b) comparing the top ten results returned from Google.ca with those of Google.cn to assess whether URLs missing from keyword search results were available elsewhere in the Google.cn database or were filtered.

The goal of this study is to assess the pervasiveness and consistency of filtering by Google on Google.cn. Using the topics of Falun Gong and Tibet, this study will assess how pervasive the filtering of each topic is by tracking the presence of the filtering statement. In addition, the study will assess how consistently keywords were filtered when searched multiple times and whether language had any effect on the consistency of filtering. Finally, the study will assess what and how much was removed due to filtering by comparing the Google.cn results with Google.ca results. The results of this study will be compared to findings of *EAIFinC* and *IFinC*. The objectives examine the degree of self-censorship exercised by Google and the content affected.

### **1.1.1 Research Questions**

This study set out to answer the following research questions from data collected during May 2008:

1. How pervasive was the filtering of search results on Google.cn?
2. How consistent was the filtering of search results on Google.cn?

3. Did language affect the frequency of filtering of searches on Google.cn?
4. When filtering occurred, what was filtered from search results on Google.cn?

### **1.1.2 Research Context**

Groups such as the ONI used Google as a tool by which they measured filtering practices around the world. The variation in accessibility from Google search results inside a country's borders (or by proxy using servers inside a country's Internet infrastructure, such as the Great Firewall of China (GFW)) and outside a country's borders have been well documented. This study was inspired by and modeled on that earlier research. The fundamental difference between this study and the prior research is its context. Previous studies focused on filtering externally imposed on Google by nation states; however, this study focuses on filtering that Google undertook internally in compliance with Chinese law. When the PRC was censoring Google, the content of search results were not affected. Access was affected: either preventing access to Google prior to searching; or, interrupting access during the execution of a search; or, preventing access to the URLs in Google's search results (OpenNet Initiative, 2006, Testing methodology chapter). With the creation of Google.cn, Google filtered the results on China's behalf. Google's decision to filter was announced publicly on their official blog and widely reported in the media at the launch of Google.cn. As Elliot Schrage, vice president of global communications and public affairs at Google, testified to at hearing of the US congress, each filtered Google.cn search was clearly labeled at the bottom of the search results page. That testimony was published immediately on Google's official blog. To be clear, the notice only informed users of the removal of results had taken place. Google's notification of its filtering did not reveal



anything about the pervasiveness or the consistency of its filtering. What was removed was a closely guarded secret (Villeneuve, 2008).

In the earlier ONI studies, an assessment of content filtering had to be based on the accessibility of websites and on HTTP header information. HTTP header “information is generally hidden from the end user” (OpenNet Initiative, 2006, p. 20). In this way Internet filtering is opaque, as studies had to test accessibility of content from Google search results; however, this study focuses on the filtering statement posted on Google.cn for the benefit of the end user. The presence of Google’s Internet filtering statement at the bottom of search results pages makes the filtering slightly less opaque, as it informs the end user filtering has taken place. Everything else about filtering remains opaque; the filtering notice is a static text that gives no details about what or how much has been filtered.

The focus of this research is entirely on Google. Earlier studies had used Google.com as a tool for tracking Internet censorship (Zittrain & Elderman, 2003; OpenNet Initiative, 2006). In this study, the focus is primarily on Google and Google’s actions, by tracking when filtering occurs and examining the specific content filtered.

The first part of this study tracks the presence of the filtering statement to analyse how pervasive the filtering was and how consistently it was taking place. The second part examines what was removed from the Google.cn search results by comparing the top ten Google.cn results with the top ten Google.ca search results. URLs present in Google.ca search results, but not appearing in Google.cn search results, were not necessarily filtered. There are a number of limitations in trying to draw clear conclusions about why a URL might not appear in Google search results. First and foremost, a URL may simply not have appeared in the top ten results but rather may have appeared elsewhere in the results list.

Second, the URL may not have been in the Google.cn database at all, or may not have been present on the server that was searched. In order to verify that a URL was actually filtered, this study searched these absent URLs again in Google.cn using a site search for the specific URL. That URL search was then analyzed for the appearance of the Google.cn filtering statement.

The Google.cn filtering statement is used to determine whether filtering of search results by Google has occurred. This is not to say that the Chinese state does not have a role in the filtering of Google.cn – it is simply not a direct role. The list of what was to be filtered was provided and the filtering was carried on in the context of a positive law: “The states that filter the Internet must choose which topics to block (the scope of filtering) and how much of each topic to filter (the depth of filtering)” (Faris & Villeneuve, 2008, p. 1). Google’s filtering of its Chinese regional search engine must be seen in this broader context. Any state that seeks to control dissidents and dissent information must have a strategy for the regulation of the Internet. For Faris & Villeneuve the control of the Internet is a natural progression from “taking over television and radio stations”; in the twenty-first century it is a “part of the everyday political and cultural reality of many states” (2008, p. 9).

## **1.2 Background**

### **1.2.1 *The Internet in China***

Since the mid-1980s, Internet access has held much promise and potential for the increased flow of information in China and thus the informing of its citizens. Many hopes (both within and without) were centred on the ability to use the Internet as a tool for the dissemination of news and ideas.

Information technology (IT) has long been hailed as the way of the future in China: “[ten] years before China’s first Internet connection, technocrats in academia and government were building support for the idea that information technology in general, and networks in particular, were essential to China’s growth” (Foster & Goodman, 2000, p. 11). The Chinese Communist Party (CCP) runs a single-party state which relies on the careful and controlled flow of state sanctioned information to maintain the stability of the country. IT would allow them to both decentralize decision making and remain in ultimate control of those decisions. The CCP saw that when “the move was made to transform the Soviet Union into a high-information, low-coercion system, the entire organizational structure unravelled. This is precisely the problem that the current leadership in Beijing is trying to grapple with” (Saich, 2004, p. 339; see also Foster & Goodman, 2000):

the advent of the Internet makes it much more difficult for the CCP to manage information flows and to ensure that its view of events prevails... [There] has been a substantial tradition in China of managing information flows to ensure that the state is the primary, if not sole, provider of information. The CCP has tried to channel information flows so that they are vertically linked and it has eschewed the horizontal flow of information.

(Saich, 2004, pp. 338, 337)

The strategy of working with Google shows the CCP’s desire to enable access to information while retaining ultimate control over what is accessible.

Economic growth must also continue to maintain the CPP’s regime. Many studies in the United States have pointed to the investment in IT and the robustness of the American economy as evidence that IT can help to foster economic growth (Foster & Goodman,

2000, p. 11). Likewise, the Chinese leadership has committed a great deal of resources to developing their own infrastructure in order to achieve similar economic results.

[The] main objectives of the Chinese leadership's regulative policy are contradictory: on the one hand, the Chinese state is attempting to exert control over Internet use for political reasons and, on the other hand, the state is attracted by the economic advantages offered by the technological modernization of information and communication.

(Damm & Thomas, 2006, p. xix)

In order to maintain their regime, the CCP must carefully regulate the flow of information; yet in order to encourage growth in the economy, and thus enhance China's status on the world stage, communications and information must be opened (Damm & Thomas, 2006, p. xix). For balance between these two to be maintained, the government must regulate Internet operations carefully and diligently.

The forbidden content in China's Internet and telecommunications laws cover a range of broad categories rather than a list of specific topics. Though topics that are sensitive are well documented and some lists of words have been discovered embedded in computer software (Qiang, 2004), there is no official or public list of banned words or topics. Rather, any information that involves the following is forbidden under Chinese law:

- (1) Contradicts the principles defined in the constitution [of the People's Republic of China].
- (2) Endangers national security, discloses state secrets, subverts the government, destroys the unity of the country.
- (3) Damages the honour and the interests of the State.

- (4) Instigates ethnic hatred or ethnic discrimination, destroys the unity of [China's] nationalities.
- (5) Has negative effects on the State's policy on religion, propagates evil cults or feudal superstition.
- (6) Disseminates rumours, disturbs social order, undermines social stability.
- (7) Spreads lewdness, pornography, gambling, violence, murder, terror or instigates crime.
- (8) Offends or defames other people, infringes upon the rights and interests of other people.
- (9) Other contents that are forbidden by law or administrative regulation. (Wacker, 2003, p. 62)

The scope of each of these categories is deliberately far reaching. The ONI identified the State Secrets Law as being particularly broad:

The State Secrets Law is a critical and broadly defined part of China's Internet content control. The State Secret Law defines the term "state secret" to include confidential information in areas ranging from social development, to technology, to international relations, to the national defence and economy. ... China has retroactively declared information a state secret. If convicted of providing state secrets to overseas individuals or Organisations via the Internet, citizens face the death penalty. (OpenNet Initiative, 2006, p. 16)

Not only are there broad regulations on forbidden content and news, but in addition, all Internet service providers and Internet cafés must be licensed and Internet users must be registered with any number of different authorities at local, provincial, or state levels depending on the specific regulation: “[in] this respect, it should be noted that Internet cafés are not only required to apply for a business licence and register with the local Public Security Bureau, but also to hire appropriate personnel to monitor the activities of users, who themselves are supposed to show an ID and register their details” (Wacker, 2003, pp. 66-67). Moreover, just as the PRC has built up its infrastructure, it has also been building up Internet regulation and the security personnel to enforce those regulation. Special police units and security groups have been created or expanded with the express purpose “of fighting cyber crime, ensuring IT security through work such as providing information and consultancy on computer viruses, and ‘keeping order’ in cyberspace” (Wacker, 2003, p. 67). The PRC has an extensive array of strategies to control citizens’ access to the Internet, from control over internet cafes to the vaguely defined but broad and powerful State Secrets Law. In working out a relationship with Google, the PRC tried to give its citizens access to the Internet while maintain control over what information could be found and connected to. At the same time working with Google allowed the PRC to portray itself in the eyes of the world as being enlightened and progressive.

### **1.2.2 Sensitive Matters: Falun Gong and Tibet**

As an objective of this study was to track the pervasiveness and consistency of filtering by Google.cn and investigate what results were being removed, the keywords searched were limited to topics known to be sensitive topics in China. Because many topics subject to Internet filtering overlap, topics of specific interest to China were chosen for this

research: Falun Gong and Tibet. These topics are taboo specifically because of how they challenge the status quo established by the CCP and the filtering of these topics online is well documented: “[in] China, Web sites that represent the Falun Gong and the Tibetan exile groups are widely blocked” (Faris & Villeneuve, 2008, p.12).

Tibet’s relationship with China is a tumultuous one dating back centuries. Tibet is currently an autonomous region of China, which is functionally similar to a Chinese province. Tibet’s struggle for true autonomy makes discussion of Tibetan politics a sensitive issue in China. The current debate dates back to 1950 when the PRC sent in troops to solidify its hold on the region. Since then there has been great upheaval and violent clashes in Tibet, particularly between those seeking independence and the PRC and its loyalists. Many sources on this topic exist. BBC’s Tibet profile provides a good overview of Tibet’s current place within China and the issues that surround that status (BBC, 2014). John and Elizabeth Roberts’ book, *Freeing Tibet: 50 years of struggle, resilience, and hope* (2009), is an in-depth source that traces the struggles of the Tibetan Independence movement beginning in 1959 with the departure of the Dalai Lama through to the book’s publication in 2009. The official source for news and information about all aspects of Tibet can be found on China Tibet Online at [eng.tibet.cn](http://eng.tibet.cn), the official English language site of the Tibetan Autonomous Region of China.

The Falun Gong is a religious movement “derived from an ancient form of *qigong*” that was found in China in 1992 by Li Hongzhi, known to its practitioners as The Master.

[Its] adherents exercise ritually to obtain mental and spiritual renewal. The teachings of Falun Gong draw from Buddhism, Daoism, Confucianism, and Chinese folklore. The movement’s sudden prominence in the late-1990s

became a concern to the Chinese government, which branded Falun Gong as a “cult.” (Introvigne, 2015, para. 1)

Initially Falun Gong was a non-issue for the PRC. It was one of many social groups dedicated to physical and spiritual improvement; Falun Gong even had the support of many officials in its early years. However as the whole *qigong* movement grew in influence the PRC sought to reign in and gain control of such groups. Falun Gong resisted such measures which lead to criticism, tensions, outlawing, and persecution. David Palmer’s book, *Qigong fever: Body, science, and utopia in China*, documents the history of *qigong* practice in 20<sup>th</sup> century China that led to the founding of Falun Gong and the events of the 1990s that led to it being denounced by the PRC. The subsequent persecution of Falun Gong is covered in detail in *Revenge of the Forbidden City: The suppression of the Falungong in China, 1999-2005* (Tong, 2009).

### **1.2.3 Google in China**

The story of Google’s development and its impact on the world has been well documented, as such, only a brief sketch will be presented here. For more complete accounts see Vise & Malseed (2005) and Battelle (2005). Google is the brain child of two Stanford University students, Sergey Brin and Larry Page. Their work was originally based on mapping the World Wide Web in mathematical terms. As the project grew Brin and Page became more interested in the ability to search and access information via the Web. In particular, they sought to associate the most relevant pages with a search. Google’s PageRank algorithm initially drove the company’s success (Vise & Malseed, 2005). As the company has grown, it has focused on Research and Development. Innovation is the



backbone of the company (Wojcicki, 2011), but Google, like any company, needs to grow through the expansion of existing markets and expansion into new markets.

China is obviously a very exciting market in general and also for Google. ... “We have actually a very significant market share in China. There’s tremendous opportunity for us there with our existing market share to make money through advertising. We’re just ramping up our business operations.” (Vise & Malseed, 2005, p. 271)

Google was facing further restrictions, the worst case scenario becoming fact: “access became slow and unreliable ... even though we weren’t doing any self-censorship our results were being filtered anyway, and our service was being actively degraded on top of that” (The Internet in China: A tool for freedom or suppression? February 15, 2006, pp. 68-69).

#### ***1.2.3.1 The Birth of Google.cn***

On January 25, 2006, Google launched a new local version of Google.cn, which “[removed] certain sensitive information from our search results” (McLaughlin, 2006, para. 2). Fully aware this was a controversial decision, Google laid out its position in two key places, first on the official Google Blog, where Andrew McLaughlin posted the official announcement about Google.cn; second, at the American congressional hearing entitled “The Internet in China: A tool for freedom or suppression?” where Elliot Schrage, VP, Global Communications and Public Affairs for Google, testified on Google’s behalf before US Congressional subcommittees. In both documents, Google’s dilemma is clearly laid out:

We faced a choice at that point: hold fast to our commitment to free speech (and risk a long-term cut-off from our Chinese users), or compromise our principles by entering the Chinese market directly and subjecting ourselves to Chinese law and regulations. (The Internet in China: A tool for freedom or suppression? 2006, p. 68)

Google decided that it would compromise its principles so that it could connect with its Chinese users and offer those users a connection to the wider Internet.

As Ang (2005) argues, Internet regulation should be in keeping with offline regulation. This point lies precisely at the heart of the Google's defence of its actions with respect to China. With the filtering of Google.cn, Google's actions in keeping with the laws of the PRC—those laws, however, are in direct contradiction with what Western countries view as limited and judicious with respect to censorship. Google, more than other companies, has opened itself to these kinds of criticisms, particularly because of its unofficial but well known corporate motto "Don't be evil". CEO Eric Schmidt acknowledged it directly at a 2006 World Economic Forum Panel discussion:

"We concluded that although we weren't wild about the restrictions, it was even worse to not try to serve those users at all," Schmidt said. "We actually did an evil scale and decided not to serve at all was worse evil," he said, referring to the company's famous "don't be evil" creed. (Cowley, 2006, para. 2)

Several months later, Google co-founder Sergey Brin indicated more clearly that the move to censor Google.cn was not precisely in keeping with Google's motto: "We felt that perhaps we could compromise our principles but provide ultimately more information for

the Chinese and be a more effective service and perhaps make more of a difference,' he said. ... 'Perhaps now the principled approach makes more sense'" (Associated Press, 2006, para. 4,7). Brin's statements were initially taken to signal a change in Google's stance towards cooperation with the CCP; however this was just as quickly and widely refuted by the company:

Brin has now reiterated Google's intention to move ahead with Google.cn, a version of the search engine that censors thousands of sites according to Chinese standards. Addressing speculation over his comments, Brin admitted to a group of invited journalists that "standing by the principle against censorship" could be interpreted as Google pulling out of China. But, he insisted, "That's an alternative path. It's not the one we've chosen to take right now." (Google denies rumor of quitting, 2006, para. 4-6)

So, Google was aware of the ethical complexity of its decision, but it decided to push ahead with its strategy of working within the censorship rules set down by the Chinese government.

Google's China problems did not end with the launch of Google.cn. The CCP continued to limit access to Google.com, which was blocked by China in May 2006; access to Google.cn remained unaffected. Yet accessing information via Google.com remained an issue as "[the] vast majority of Web users inside China – 99 percent [Brin] said – use Google.com rather than Google.cn at this time" (Rothstein, 2006, .Com or .Cn? section para.1).

Google Inc. remained strong in its defence of its Chinese policy decisions. At the first Internet Governance Forum, a United Nations special conference, Google

characterized itself as a champion of minority languages and rights and valiant against unfair copyright laws.

The Silicon Valley giant [attempted] to position itself as a force for change that can finance web entrepreneurs in the developing world, champion the rights of consumers against “over-zealous” copy-right laws and use the web to protect diverse minority cultures and languages...[declared] itself unrepentant over the controversial decision to censor its search engine at the behest of Beijing...the firm [insisted] its presence in China does more good than harm by getting more information to more people. (Smith, 2006, para. 2-3)

#### **1.2.3.2 Pulling out of China**

While there was no interference with access to Google.cn, in January 2010 Google revealed publicly that it had been the subject of various cyberattacks in China. This was an unusual move for the company, which by its own admission, was not normally so public about such incidents (Drummond, 2010a).

Google traced the attacks to China and although the attribution regarding the Chinese government is unclear, Google also discovered that the attackers also attempted to compromise the Gmail accounts of Chinese human rights activists. (Villeneuve, 2010, para. 1)

During its investigation Google discovered it was only one of at least twenty companies that were deliberately targeted with the goal of limiting free speech. These attacks and its motives prompted Google to reconsider activities on Google.cn:

We have decided we are no longer willing to continue censoring our results on Google.cn, and so over the next few weeks we will be discussing with the Chinese government the basis on which we could operate an unfiltered search engine within the law, if at all. We recognize that this may well mean having to shut down Google.cn, and potentially our offices in China. (Drummond, 2010a, para. 8)

Three months later, on March 30, 2010 Google stopped censoring Google.cn and redirected users arriving there to Google Hong Kong (Google.com.hk), “where [they offer] uncensored search in simplified Chinese, specifically designed for users in mainland China and delivered via [their] servers in Hong Kong” (Drummond, 2010b, para.2). Google maintained a business presence in China, though it was small and comprised mostly of sale and research and development. Less than a month later Google launched what has become known as the Transparency Report also developed and implemented “to show how laws and policies affect access to information online, including law enforcement orders for user data and government requests to remove information” (Callaghan, 2014, para. 1). Since then Google has expanded this service as it strives to maintain its corporate integrity.

## Chapter 2 Literature Review

### 2.1 Removing Sensitive Materials from the Internet

Though often heralded for the possibility and potential it holds for greater distribution of and access to information, the infrastructure of the Internet—through both hardware and software—can be used to reduce and censor that flow of information. Likewise, the threat of censorship can also create a culture of self-censorship.

Ang offers a thoughtful and well-researched approach to Internet regulation: “[The Internet] highlights a major legal issue of global interconnectivity: which censorship standard should be applied? ... Material that is illegal in one country and punishable with a heavy sentence may be wholly legal in another” (2005, p. 46). Using specific cases of censorship and content regulation, Ang examines the history of Internet regulation, problems, methods, and censorship trends. Moreover, he cuts to the heart of the censorship debate in addressing the goals of censorship itself.

Censorship “is not meant to be one hundred percent effective” (Ang, 2005, p. 40). The notion that censorship should be one hundred percent effective and worthless if it is not is fallacious:

This is a form of what economists call the “Nirvana fallacy,” i.e., something is not worth doing if it does not achieve perfect results. An argument often made against censorship is that on the Internet, a user can always make an international phone call anyway. The argument misses the point. It has never been possible to reliably block everything, even at customs. The persistent will always find a way past the censors. The censor’s goal is

achieved if it is difficult for *most* users *most* of the time to access *most* of the material. (Ang, 2005, p. 40)

The problem of trying to censor the Internet arises from both the infrastructure itself and from the numerous methods of communication which depend on that infrastructure: “one-to-one (email) to one-to-many (websites) and many-to-many (Usenet groups)” (Ang, 2005, p. 40). As noted in the section 2.2 Internet Filtering in China below, this strategy of blocking access to “*most* of the users *most* of the time to *most* of the material” is crucial to the CCP’s censorship policies and strategies (Ang, 2005, p. 40).

The ONI is an important source of information about Internet censorship and filtering practices around the world, and for identifying and defining the technical methods (such as technical blocking, removal of search results and take-down) and socio-political and legal strategies employed which lead to peer-pressure and self-censorship (OpenNet Initiative, n.d.-a). Each of these methods has its own uses and reasons for application. Likewise, there may be multiple reasons for the application of any one of these filtering mechanisms. For instance the ONI points out that filtering at an institutional level may be done because of governmental regulations or at the request of the government, but more often it is done “to meet the internal objectives of the institution such as preventing the recreational use of workplace computers” (OpenNet Initiative, n.d.-a, Institutions section, para.1).

Internet regulation is a basic question of governance for every nation. In his editorial, “China-bashing in the Internet censorship wars” for *Online Information Review*, Gorman deliberately uses an America institutional example of Internet censorship:

The simplest and best examples come from the USA ... “even as the US condemns Internet censorship abroad, our government limits the content available to citizens here at home. The CIPA [Children’s Internet Protection Act] requires libraries and schools to filter Internet content in order to qualify for federal funds to help pay for computers and Internet access”. (2005, pp. 453-454)

Gorman is quick to point out that “[the] USA, Britain, Australia, New Zealand, Canada, and members of the EU all censor the Internet on some level. ... [In his] opinion, any sensible view of the [Internet] must admit that some sort of censorship or regulation is necessary, and this is put into practice differently by different societies” (2005, pp. 453, 455).

One of the instances of content regulation Ang brings out is a case involving Yahoo and the auction of Nazi memorabilia in France, which has strict laws against Nazism:

A French court in 2000 ordered the Internet search engine Yahoo to block French users from accessing a section of the site that auctioned Nazi memorabilia. Although Yahoo initially objected, it later banned the sale of Nazi and hate-related material on its site. ... First, Yahoo had used a French domain name yahoo.fr. That gave the French court a toehold to assert its jurisdiction on Yahoo. Second, Yahoo was targeting the French users through its advertising. (2005, p. 40)

Removing information is just one of the ways in which companies that provide services online regulate content based on national laws.



Lumen, previously known as Chilling Effects, “includes complaints of all varieties, including those concerning trademark, defamation, and privacy, both domestic and international” (Lumen, 2015, para. 3). These requests for removal of online information generally relate to racist, pro-Nazi or holocaust denying material. For example in the fall of 2004 the Canadian Jewish Congress demanded the removal of Zundelsite, a website in support of German-Canadian Holocaust denier and Nazi supporter Ernst Zundel (Lumen, n.d.-a). There were also a number of complaints requesting the removal of racist and pro-Nazi material for France and the removal of Holocaust deniers posting in the German regional section of Google Groups (Lumen, n.d.-b; Lumen, n.d.-c). These complaints all point out specifically that such material is punishable by law in the respective country.

Censorship applied in a limited and judicious manner is not inherently ‘evil’ *per se*. Ang lays out clearly how censorship should be applied:

This is the basic approach that governments should take: one should not gain nor lose rights merely by going online. So it should not be surprising if Germany attempts to outlaw anti-Semitic sites hosted in Germany because that is the law for offline media. (Ang, 2005, p. 5)

Internet regulation and censorship should not be disproportionately harsh or lax. Rather such controls should be in keeping with the laws and regulations of a country, maintaining the rights and responsibilities that exist offline.

## **2.2 Internet Filtering in China**

Controlling the Internet is not simply a legal matter. The ONI has also identified China’s extra-legal content controls placed on both Internet service providers and users:

[The] Internet Society of China pressures content and access providers to agree to a “Public Pledge of Self-Regulation and Professional Ethics.” Companies often accede.... Internet regulation in China is based on the philosophy that “one is responsible for what one publishes.” Thus, Internet companies in China practice a high degree of self-censorship. (OpenNet Initiative, 2006, p. 18)

Moreover, users themselves who are in agreement with Chinese content regulations will take the philosophy of personal responsibility a step farther, not only self-policing but policing the content of others as well: “[some] citizens view Internet regulation as necessary, and monitor Web sites, chat rooms, and bulletin boards for inappropriate content, reporting violations to authorities” (OpenNet Initiative, 2006, p. 18). This problem is further compounded by the fact that what exactly constitutes sensitive material is not made explicit to the users and the notices to Internet service providers are in no way meant to be exhaustive. Power resides in implicating the user in the sociocultural dynamics of censorship.

Should a user try to access sensitive content, any number of different strategies could be employed by the government of the PRC to block access. In the March 2008 issue of *Atlantic Monthly*, Fallows examined China’s content filtering strategies: “[in] the few seconds after a user enters a request at the browser, and before something new shows up on the screen, at least four things can go wrong—or be made to go wrong. The first and bluntest is the ‘DNS block’” where the Domain Name System, or the actual numerical address associated with a webpage, “is instructed to give back no address, or a bad address, [with the result that] the user can’t reach the site in question” (Fallows, 2008, p. 66).

Second, the connection could be reset by the Chinese servers while trying to connect to a blacklisted site. In this case both the user's computer and the webpage will be told the connection has been reset, but not why (e.g., because the page has been blacklisted). Third, URLs may be filtered and blocked based on keywords in the URL, such as Falungong.com, which contains blacklisted keywords in the URL. "The forbidden list contains words in English, Chinese, and other languages, and is frequently revised" (Fallows, 2008, p. 67). Fourth, the actual content of the page may be scanned "to judge its page-by-page acceptability" (Fallows, 2008, p. 67).

The CCP content regulations are *not* meant to be exhaustive, indeed such censorship strategies are *not* meant to be infallible. They are meant to create and breed a culture of self-censorship, so that "the philosophy that 'one is responsible for what one publishes'" is a way of life not just theory (OpenNet Initiative, 2006, p. 18). Saich points to this as the possible outcome:

The Chinese leadership is clearly aware that it cannot completely control the flow of information or access to forbidden sites by its citizens. Its intention is to lay down warnings about the limits of the permissible and to deter the casual browser from becoming too inquisitive about the world outside. In this limited respect, it may be successful. (Saich, 2004, p. 340)

Since then the notion that self-censorship is effective in China has only been borne out with increasing evidence. Fallows is very clear that "[what] the government cares about is making the quest for information just enough of a nuisance that people generally won't bother" (Fallows, 2008, p. 69). ConceptDoppler, an Internet filtering research project which is discussed in more detail in the next section, has drawn this same conclusion from

their research: “[our] Internet measurements suggest that the Great Firewall of China’s (GFC) keyword filtering is more a panopticon than a firewall, i.e., it need not block every illicit word, but only enough to promote self-censorship” (Crandall, Zinn, Byrd, Barr & East, 2007, p. 1).

The panopticon is a particularly apt metaphor for the regulation of the Internet in China. It was first introduced a concept for the Internet in China by Tsui in his 2001 Master’s thesis (2001). Tsui later expanded on the idea:

The Panopticon, invented by English philosopher Jeremy Bentham (1791) and mediated by Michel Foucault, is a prison where the inmates are being watched by an invisible guard. The uncertainty of being watched regulates and normalizes the inmates’ [behaviour]. In this article, panopticism is used to show how the Internet in China is regulated by way of surveillance. (Tsui, 2003, p. 66)

The fact that the content *regulation* and the *penalties* are known, but that the forbidden *content* is non-specific and subject to change, possibly retroactively, has a fundamental impact on creating and maintaining a culture of self-censorship. This ambiguity is fundamental to the overall impact of content filtering in China and Internet culture within China, according to the ONI:

Importantly, China’s filtering efforts lack transparency: the state does not generally admit to censoring Internet content, and consequently there is no list of banned sites and no ability for citizens to request reconsideration of blocking.... The topics defined as sensitive, or prohibited, by China’s legal code are broad and non-specific, and enforcement of laws such as the ban

on spreading state secrets discourages citizens from testing the boundaries of these areas. China's legal and technological systems combine to form a broad, potent, and effective means of controlling the information that Chinese users can see and share on the Internet. (OpenNet Initiative, 2006, p. 52)

All of this serves to reinforce self-censorship by the users and creates an acceptance of filtering practices among Internet users as a whole. Harwit and Clark (2006) argue that:

Internet demographics indicate that user profiles, in fact, work toward social stability, at least in the coming few years. Younger users may avoid controversy, and use the network more for education and entertainment purposes. ... [As] today's elite Internet users age and perhaps encounter financial or political difficulties in their lives, and as more disaffected members of society find access to the network, we may see it emerge as a tool to be utilized more frequently to channel discontent. ... [Because] the nature of the Internet's audience means that the Internet will remain an unlikely tool for precipitating socially disruptive forces. (2001, pp. 36-37)

Ultimately the attitude of the Chinese Internet users proves that the PRC's censorship strategies do not need to be exhaustive to be effective. By limiting users' abilities to publish and to access materials deemed sensitive by the state, the attitude in China has begun to move past a simple culture of fear or coercion, or even one of acceptance. In fact, the Chinese government has been able to create a culture among citizens in China that not only accepts censorship but engages in self-censorship and in the policing of others on the Internet.

### 2.3 Circumventing the Censors

Not all Chinese citizens are ready or willing to accept the limits placed on them by the PRC. Numerous methods have been developed and deployed to subvert and circumvent the Internet censors, regulations and bypass the technological barriers. The Citizen Lab (2007) at the University of Toronto created a guide that summarizes methods of Internet censorship.

An important aspect of this subversion is the use of language to circumvent the keyword censorship. Methods regularly employed include using alternate languages to search or access materials (Citizen Lab, 2007, p. 28), adding symbols or puns to modify words, substituting homophones (i.e., swapping one Chinese character for another that sounds the same but is written differently) (Wiener, 2011; Qiang, 2011). ConceptDoppler (Crandall et al., 2007) offers a “Discussion of Keyword-Based Evasion” that is more technical than the other articles (notably inserting HTML comments, IP packet fragmentation, and using different encodings amongst other techniques).

The most famous of this language play is the Grass Mud Horse which has become a rallying cry for those fighting back against the censorship of the PRC:

The Grass Mud Horse—or *cao ni ma*, the homophone of a profane Chinese expression—became the de facto mascot of Chinese netizens fighting for free expression. It inspired poetry, videos, and clothing lines. As one blogger explained, the Grass Mud Horse represented information and ideas that could not be expressed in mainstream discourse. (Qiang, 2011, p. 52)

The true success of this meme is evident in the fact it has been taken up and used in many different media and locations online and that it is even possible to purchase merchandise

depicting the Grass Mud Horse. Of course, the Grass Mud Horse is only the most famous example; this practice of modifying language is examined in detail in the article, “Grass-mud horses to victory: The phonological constraints of subversive puns” (Wiener, 2011).

Another relatively low-tech method for subverting the Chinese censors is using of mirror sites and alternate domain names.

One of the most common ways to censor a website is to block access to its domain name, e.g. news.bbc.co.uk. However, sites are often accessible at other domain names such as newsrss.bbc.co.uk. Therefore if one domain name is blocked try to see if the content can be accessed at another domain.

Example: news.bbc.co.uk → newsrss.bbc.co.uk. (Citizen Lab, 2007, p. 28)

Likewise, news articles and information are reposted “inside the country by a small but active group of tech-savvy “information brokers” who know how to circumvent the Great Firewall and circulate the news via BBSs, mass e-mailings, and other online channels” are another essential source of information to those within China whose Internet reach is stunted by the GFW (Qiang, 2011, p. 53).

## **2.4 Researching Internet Filtering in China**

Many studies have examined all aspects of Internet censorship in China, including: rates of filtering, effected content, methods of execution, involvement of various companies, and impact on Chinese Internet users, particularly academics and other researchers. As the focus of this study is the filtering practices of Google.cn, the primary focus of this section literature review will focus on the research up to 2010 when Google ended Google.cn.

A comprehensive meta-analysis of the research from the last 20 years was recently done by Herold and de Sela (2015). Among the most recent and ambitious studies, their article reviewed all types of publications on Internet filtering in China from 1990 to early 2013. This article provides the most complete overview of the research to date and the most comprehensive reference list. Herold and de Sela review notes important issues with the research as a whole. First, they question the way researchers engage the existing literature. They note a lack of cross-discipline engagement as many research only engage with literature within their own discipline and identify a trend of ignoring research beyond five years old. They find these things troubling as it leads to both weaknesses and duplication in the overall body of research. Second and most important, Herold and de Sela question how Chinese Internet research is done and “whether this research are actually exists” (Herold and de Sela, 2015, p. 78). The authors conclude that the Chinese Internet research needs a more systematic approach and more focus on the Chinese citizens’ Internet practices. Finally, “the authors ... point out that national Internets should not be essentialized” (Herold and de Sela, 2015, p. 78). They raise deep concerns about the divide between western researchers and the research subject, the biases of western researchers and the trend of ignoring research by Chinese researchers because it is written in Chinese.

Amongst the many studies are two important benchmarks: *Empirical analysis of Internet filtering in China (EAIFinC)* and *Internet filtering in China in 2004-2005: A country study (IFinC)* (Zittrain & Elderman, 2003; OpenNet Initiative, 2006).

*EAIFinC* was a six-month long study in 2002 that identified four distinct methods of filtering and documented a marked change in the sophistication of the filtering part way through the study:



The authors [concluded] (1) that the Chinese government maintains an active interest in preventing users from viewing certain web content, both sexually explicit and non-sexually explicit; (2) that it has managed to configure overlapping nationwide systems to effectively—if at times irregularly—block such content from users who do not regularly seek to circumvent such blocking; and (3) that such blocking systems are becoming more refined even as they are likely more labor- and technology-intensive to maintain than cruder predecessors. (Zittrain & Elderman, 2003, para. 1)

The *EAIFinC* was subsumed into the larger OpenNet Initiative, which produced the larger follow up study *IFinC*, which is one of in a suite of studies that the ONI conducts to monitor Internet filtering around the world. The ONI study, *IFinC*, set out to:

determine the type, location and behaviour of [Chinese] filtering technology.... For testing, depending upon a series of local factors, ONI [obtained] network access at multiple levels through a combination of: Proxy servers, Long distance dial-up, Distributed application, and Dedicated servers. (OpenNet Initiative, 2006, pp. 19-20)

A number of different aspects of Internet content filtering were examined including: email filtering, blog filtering, Google cache testing, and filtering by Chinese search engines Baidu and Yisou. Differences between proxy testing and in-state testing of domains and URLs were compared. The ‘long list’ testing from *IFinC* was the original model for this study. The long list was “a testing list ... containing the top 10 sites returned by the Google search engine for queries on Chinese and English keywords related to sensitive topics, such as ‘Falun Gong’ ... the results of in-state testing of this long list [were reported] by topic area”

(OpenNet Initiative, 2006, p. 28). The *IFinC* study also includes longitudinal comparison of “proportion of top 10 and top 100 Google search results inaccessible” of the 2002 and 2005 results for a dozen search terms (OpenNet Initiative, 2006, p. 36).

#### **2.4.1 *The Google.cn Years 2006-2010***

Google’s announcement about Google.cn in January 2006 set off a flurry of research and activity in response. The United States Congress, Amnesty International and Human Rights Watch all addressed the impact of American tech companies’ decisions to engage in Internet filtering in accordance with the laws of the PRC in 2006. Each of these reports deals with the impact and the ethics of the choice made by Google Inc., among others, to filter their search results on behalf of the PRC (The Internet in China, 2006; Amnesty International, 2006; Human Rights Watch, 2006)

The Subcommittee on Africa, Global Human Rights, and International Operations and the Subcommittee on Asia and the Pacific of the Committee on International Relations, House of Representatives, United States Congress, held hearings entitled “The Internet in China: A tool for freedom or suppression?” in February, 2006. The report of the subcommittee is worth attention because a number of Google’s executive officers testified at the hearing. As Elliot Schrage, vice president of global communications and public affairs at Google, testified that Google’s goal was to satisfy:

- (a) First, our business commitment to satisfy the **interests of users**, and by doing so to build a leading company in a highly competitive industry; and
- (b) Second, our policy conviction that **expanding access** to information to anyone who wants it will make our world a better, more informed, and freer place. (Schrage, 2006, Introduction section, para. 6-7)

While at the same time, being “responsive to **local conditions**” like those in China (Schrage, 2006, Introduction section, para. 9). According to Schrage, notifying users when results were removed, maintain users’ privacy, and continued access to Google.com unfiltered Chinese language search balanced these concerns.

In July, Amnesty International published *Undermining freedom of expression in China: The role of Yahoo!, Microsoft and Google* addressing the impact on the freedom of expression in China framing their analysis of corporate responsibility within the “Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights and other international human rights treaties” (Amnesty International, 2006, p. 7). Interestingly, the authors’ primary analysis is of each of the company’s own documents relating to their values and mission statements and compares these statements to the actions they have taken with respect to China.

Human Rights Watch’s “*Race to the bottom*”: *Corporate complicity in Chinese Internet censorship* was published later that summer and expands on the themes laid out by Amnesty International. They expand their work to include quantitative analysis of “URL de-listing on Google.cn, Yahoo! China, MSN Chinese and Baidu” and “Comparative keyword searches on Google.cn, Yahoo! China, MSN China, Baidu, Yahoo.com, MSN search and Google.com” (Human Rights Watch, 2006, pp. 142-145). This report also stands apart from the rest of the literature as it makes concrete recommendations to various groups with vested interests in Internet censorship in China, including: 1) the PRC, 2) Internet companies, 3) their investors, 4) home nations of these Internet companies, 5) international governmental agencies, such as the UN and the World Trade Organization,

6) activist and nongovernmental organizations, human rights groups, etc., and 7) internet users.

Leading up to and during the Beijing Summer Olympics in 2008, the ONI reported regularly on its blog about the impact of the Olympics on China's Internet filtering. While on August 1, they reported freer access to sensitive materials, it was unclear how widespread this access was (OpenNet Initiative, 2008a). This reporting was followed up by "[comparing] data from the Olympics Main Press Center (MPC) to that from other locations in Beijing, compiling a snapshot of Internet filtering in China leading up to week 1 of the Olympics" (OpenNet Initiative, 2008d, para. 1). By end of the year, it was reported "the Chinese government [had] begun to reinvigorate its filtration of foreign websites, including bans on BBC's Chinese language website and Voice of America in Chinese" (OpenNet Initiative, 2008b, para. 1).

In 2008, the ONI's work on China was put in a broader context of worldwide Internet filtering in the first chapter of *Access denied: The practice and policy of global Internet filtering* (Faris & Villeneuve, 2008). Faris and Villeneuve analysed the ONI 2006 findings from 40 countries to compare aspects of filtering such as scope, amount, blocking techniques, consistency and transparency/accountability. China was found to be one of the most rigorous Internet filtering regimes. They filtered a large amount of a wide range of topics, used a number of different techniques to filter material and low transparency/accountability. While tracking and measuring Internet filtering can be difficult it, is possible, however researching and establishing the motivations and impacts of that filtering are more difficult. Faris and Villeneuve articulate the inherent difficulty of Internet censorship research:

Measuring and describing Internet filtering defies simple metrics. Ideally, we would like to know how Internet censorship reduces the availability of information, how it hampers the development of online communities, and how it inhibits the ability of civic groups to monitor and report on the activities of the government, as these answers impact governance and ultimately economic growth. However, this is much easier to conceptualize at an abstract level than to measure empirically. Even if we were able to identify all the Web sites that have been put out of reach due to government action, the impact of blocking access to each Web site is far from obvious, particularly in this networked world where information has a habit of propagating itself and reappearing in multiple locations. (Faris & Villeneuve, 2008, p. 11)

Villeneuve also produced his own study through the Citizen's Lab at the University of Toronto. Villeneuve's study is of particular note because it is one of the few academic studies that studied Google.cn while it was being filtered by Google. The focus of this research was to “[interrogate] and [compare] the censorship practices of the search engines provided by Google, Microsoft and Yahoo! for the Chinese market along with the domestic Chinese search engine Baidu” (Villeneuve, 2008, p. 1). The Villeneuve study focused on the transparency and extent of each company's filtering practices. The transparency of all search engine providers was found to be low, though Google was found to censor the least of the companies tested (Villeneuve, 2008, p. 19). Ultimately the author concluded “that independent monitoring is required to evaluate their compliance with public pledges regarding commitments to transparency and human rights” (Villeneuve, 2008, p. 1).

When Google Inc. announced in January 2010 that it might have to shut down its Chinese operations, the journal *Nature* commissioned a survey of 784 scientists “about how they use Google’s products, and how integral it is to their research” to find out how they would be impacted (Qiu, 2010, p. 1012). The survey found that Google was the primary search engine used by those surveyed and it was used heavily for academic research and scientific news.

#### **2.4.2 Overblocking**

In addition to the filtering of sensitive topics, research has shown that materials not considered sensitive can be caught by filters and blocked as well. Known as overblocking, the blocking of unrelated content that is similar to sensitive material has been found in many of the studies on Internet filtering in China (e.g., Zittrain & Elderman, 2003, OpenNet Initiative, 2006, Crandall et al., 2007, Villeneuve, 2008). *IFinC* concluded that China had the most sophisticated and dynamic filtering system of any country, yet:

We documented instances of overblocking, where sites with superficial similarities to those with sensitive material, but different content, were filtered. This likely indicates China’s willingness to tolerate blocking unrelated content to prevent access to sensitive materials. (OpenNet Initiative, 2006, p. 23)

The most targeted research project on overblocking was done by the research group known as ConceptDoppler (Crandall et al., 2007). The focus of the ConceptDoppler project was to examine the more technical aspects of a) how and b) to what extent the GFC is filtering by keywords. Their work, they are quick to point out, could be used by those on either side of the Internet censorship debate:

The ability to filter keywords is an effective tool for governments that censor the Internet. Numerous techniques comprise censorship, including IP address blocking, DNS redirection, and a myriad of legal restrictions, but the ability to filter keywords in URL requests or HTML responses allows a high granularity of control that achieves the censor's goals with low cost. (Crandall et al., 2007, p. 2)

The team at ConceptDoppler discovered that 83.3% of all filtering in their testing was done by China's largest ISP, ChinaNet, and that "99.1% of all filtering that occurred at the first hop past the Chinese border" (Crandall et al., 2007, p. 1). Crandall et al.'s finding is important because it indicates that filtering is not happening strictly at the border with the GFW as generally believed or assumed. The filtering practices are more robust and incorporate many more layers of sophistication.

## **2.5 Conclusion**

While there has been much research done on many aspects of Internet filtering in China, including Google's participation and decision to engage in this filtering, Google's filtering practices on Google.cn remain relatively unknown. The aim of this study, therefore, is to build on the existing research by answering the research questions:

1. How pervasive was the filtering of search results on Google.cn?
2. How consistent was the filtering of search results on Google.cn?
3. Did language affect the frequency of filtering of searches on Google.cn?
4. When filtering occurred, what was filtered from search results on Google.cn?

## **Chapter 3     Methodology**

This is a quantitative research study that adapts a methodology used in the earlier work by the Open Net Initiative (2006) and Zittrain and Elderman (2003). Their methodology, based primarily on tracking URLs, was adapted in the current study to track keyword search filtering using a known keyword list, from which the list of filtered URLs is compiled. This change was made because Google.cn filtering affected the content of the search results displayed to the user while China's methods were not able to intervene in this way.

### **3.1     Keyword Selection**

Initially, a list of over 1000 keywords from five sources relating to sensitive materials was compiled for this study (List of words censored by search engines in Mainland China, 2006; OpenNet Initiative, 2006; Qiang, 2004; Washington Post, 2006; Zittrain & Elderman, 2003). Although a program could have been written to collect all the data required, Google's Terms of Service do not allow any automated usage of Google's services, such as computer programmes performing Google Searches, (Google, 2007), therefore all Google.cn and Google.ca searches needed to be executed manually. Following an evaluation of the time requirements to collect data on such an ambitious list of keywords, the choice was made to limit the keywords to two topics and collect only the top ten search results. Because of their highly sensitive nature in the PRC, Falun Gong and Tibet were chosen as topics. The list of keywords was trimmed to only terms relating to Falun Gong and Tibet leaving a subset of 180 keywords. The keywords, split between Chinese and English, were selected specifically to probe Google's filtering practices (see Table 1).



Table 1 Keyword counts by topic and language

Language	Topic		Total
	Falun Gong	Tibet	
Chinese	78	17	95
English	67	18	85
Total	145	35	180

The Google.cn top ten keyword search results were then compared to Google.ca top ten keyword search results to compile a list of URLs. These URLs were then searched again in Google.cn using a “site:” search operator which restrict the results to the specific site or domain (Google, 2015).

### 3.1.1 Variables

This study contains five variables: *topic*, *language*, *run*, *filtered*, and *URLs*. These variables were examined with respect to two data sets, the keyword search data and the URL site search data.

Table 2 Variable descriptions

Variable	Description
<b>Topic</b>	whether the keyword relates to 1) Falun Gong or 2) Tibet
<b>Language</b>	whether the keyword is 1) Chinese or 2) English
<b>Run</b>	single, complete execution of all 180 keyword searches
<b>Filtered</b>	whether or not the filtering statement appeared (as a binary value for calculations, 0=not filtered 1=filtered)
<b>URLs</b>	site found in the Google.ca top ten keyword search results but not present in the Google.cn top ten keyword search results

## 3.2 Data Collection Instrument

The collected data (page source code from top ten search results from Google.ca and Google.cn) was stored using a custom written MySQL database. This database

contained, along with page source code, the list of study keywords. A custom website was created for data collection. Both the database and web tool were created by local Halifax programmers Shamus Peveril and Gavin Anderegg of North Knight Studios. The web tool and the database were hosted on a server provided courtesy of the Dalhousie Faculty of Computer Science.

The web tool was divided into 4 sections.

Section 1: The first section was the login (see Figure 2). Research assistants (RAs) accessed the study's web tool with a user-chosen login and were assigned a generic ID number by the database in order to identify them.

Figure 2 Data collection web tool: Section 1 login screen



The image shows a screenshot of a web form. At the top, there is a label "Enter your email address:" in a small, dark font. Below the label is a single-line text input field. Underneath the input field is a rectangular button with the word "Submit" written on it in a simple, sans-serif font. The entire form is enclosed in a dark green border.

Section 2: The second section was the first part of data collection. A keyword was assigned to that user so data for that keyword in that search run would not be duplicated. Labeled as Step 1 of 3, the keyword was presented in the form of a Google.ca search result URL so that the RA could copy the entire URL directly into the address bar of a browser window (see Figure 3). RAs were then asked to copy and paste the source code of the search results into a text box on the same page. Also included in the Step 1 of 3 page with the data collection instructions, was:

- a link to information about finding and copying source code,
- a sample of Chinese text with links to language packs in this notice:
  - If you are seeing question marks, boxes, or another place holder, instead of Chinese characters here: “当地法律法规和政”, please install *both* Simplified [hyperlinked to Microsoft support site] and Traditional [hyperlinked to Microsoft support site] Chinese Language Packs for Windows XP.
- And, the following notice with procedures in the event of query timeout:
  - Note: If the query times out, please **email us** [hyperlinked to study email] noting your keyword, the date, and the time of day. Please log out to return this keyword to the queue and log in again if you would like to continue with a new keyword.

Figure 3 Data collection web tool: Section 2 Google.ca keyword search

Not keturah.jane@gmail.com? [Click here to log out.](#)

**Your keyword:** 法悞

**Step 1 of 3**  
Please copy the following URL into a browser:

```
http://www.google.ca/search?q=法悞&ie=UTF-8&oe=UTF-8&num=10
```

Copy and paste the source code into the text area below. [Click here for more information.](#)

If you are seeing question marks, boxes or another place holder, instead of Chinese characters here: "当地法律法规和政", please install *both* [Simplified](#) and [Traditional](#) Chinese Language Packs for Windows XP.

**Note:** If the query times out, please [email us](#) noting your keyword, the date, and the time of day. Please log out to return this keyword to the queue and log in again if you would like to continue with a new keyword.

Section 3: The third section of the web tool, labeled Step 2 of 3, had two differences from the Section 2 (see Figure 4). RAs were given a Google.cn search result URL (rather than Google.ca) and the notice about installing language pack was removed.

Figure 4 Data collection web tool: Section 3 Google.cn keyword search

Not keturah.jane@gmail.com? [Click here to log out.](#)

**Your keyword:** 法楞

**Step 2 of 3**  
Please copy the following URL into a browser:

```
http://www.google.cn/search?q=法楞&ie=UTF-8&oe=UTF-8&num=10
```

Copy and paste the source code into the text area below. [Click here for more information.](#)

**Note:** If the query times out, please [email us](#) noting your keyword, the date, and the time of day. Please log out to return this keyword to the queue and log in again if you would like to continue with a new keyword.

Once the RA submitted the source code in Step 2, the database analysed the Google.ca and Google.cn source code to compare the URLs returned.

Section 4: In the fourth section, labeled Step 3 of 3, RAs were given a Google.cn site search result URL for each of the URLs that was in the Google.ca top ten results but not the Google.cn top ten results (see Figure 5). Each of the automatically generated URLs were paired with a text box for the source code of that search to be pasted. All the same notices from the previous section appeared again with one additional note:

- If you find any anomalies, errors, other unique or interesting results, that may be of interest in this study, please [email us](#) [hyperlinked to study email] a screen shot of the Google China result page. [Click here for information on how to take and email screen shots.](#) [hyperlinked to: <http://www.wikihow.com/Take-a-Screenshot-in-Microsoft-Windows>]

Figure 5 Data collection web tool: Section 4 Google.cn URL site search(s)

Not keturah.jane@gmail.com? [Click here to log out.](#)

**Your keyword:** 法榜

**Step 3 of 3**  
Please copy each URL into a browser (double- or triple-click to select the entire URL).  
Copy and paste the source code into the text area below. [Click here for more information.](#)

**Note:** If the query times out, please [email us](#) noting your keyword, the date, and the time of day.  
Please log out to return this keyword to the queue and log in again if you would like to continue with a new keyword.

If you find any anomalies, errors, other unique or interesting results, that may be of interest in this study, please [email us](#) a screen shot of the Google result page. [Click here for information on how to take and email screen shots.](#)

Upon clicking the submit button on Step 3, RAs were thanked and then given the option of logging off or retrieving another keyword.

### **3.3 Procedure**

#### **3.3.1 *Pilot Study***

A pilot study was conducted in April 2008. The pilot consisted of one full run of the 180 keywords. The purpose of the pilot was to assess all the data collection instruments and instructions; the pilot included beta testing of the website and the mySQL data collection database and review of all instructions to research assistants (RAs). During beta testing, particular attention was paid to multi-lingual characters. Minor changes were made to the instructions; information regarding taking screenshots and installing Traditional and Simplified Chinese language packs was clarified and expanded to assume the RAs only had a basic level of computer knowledge. The data collection tool was reprogrammed to deal with user timeout, before starting the pilot there was no dealing with a keyword that was assigned to a user but not data collection was not completed.

#### **3.3.2 *Volunteer research assistants***

Volunteer RAs were recruited by personal e-mail in order to collect the data. In all 30 RAs assisted with data collection, though the majority of data was collected by six individuals. The majority of RAs were graduate or undergraduate students at a Halifax, NS university; the rest were members of the general public. They were not compensated financially, and no financial compensation was offered to anyone involved in this study. No specialised skills were required and no training was given. All instructions to the RAs was included in the web tool.

### **3.3.3 Data Collection**

Using the study web tool RAs were provided with instructions, custom Google URLs, and links to language plugins necessary to fully participate in data collection. This eliminated any need for the RAs to have any knowledge of Chinese. The web tool was available during the four weeks of the data collection period. RAs could log in and collect data at their own discretion.

Data collection was performed in runs, where one run was the complete data collection for all 180 keywords. So that all keywords were searched before beginning to search any keyword an additional time. In all, seven runs of the keywords were completed in the time allotted. In other words, data was collected seven times for each keyword over the course of the data collection period in May 2008.

For each run, one or more users were provided with a keyword to search. Upon completion of all related searches, the keyword was marked as complete for that run. Once all keywords were marked complete, the application began the next run. If a user did not complete all the necessary searches for a keyword, the keyword would time out after 30 minutes of inactivity by the user and be repopulated at the bottom of the keyword list for that run.

Data collection, including the pilot, followed a 14-step procedure which is outlined below. The logic of the process was integral to the development of the data collection web tool as it ensured consistency in data collection:

- Step 1. Keywords and associated metadata were loaded into a database by North Knight Studios programmers



- Step 2. At the beginning of each run, the keywords were sorted into a random order
- Step 3. An RA logged into data collection website
- Step 4. A keyword was selected from the randomized list and given to the RAs via the data collection web tool
- Step 5. Using an individual keyword, the RA queried Google.ca
- Step 6. The RA retrieved the page source for top ten results and pasted it into the web tool
- Step 7. Using the same keyword, the RA queried Google.cn
- Step 8. The RA retrieved the page source for top ten results and pasted it into the web tool
- Step 9. The web application compared the Google.ca top ten search results with the Google.cn top ten search results. URLs that appeared in both results for the search term were logged in the database. URLs missing from the Google.cn search results were compiled into a list and presented to the RAs.
- Step 10. Using a given URL identified in step 9, the RA searched Google.cn
- Step 11. The RA retrieved the page source for the top ten results and pasted it into the web tool
- Step 12. The RA returned to step 9 and repeated step 10 and 11 for each of the missing URLs identified in step 9
- Step 13. Steps 3 to 12 were repeated until all of the keywords were searched, completing a run, the database automatically returned to step 2

Step 14. Steps 2 to 13 were repeated for four weeks

### **3.4 Data Analysis**

The data gathered from the Google search results were collated and stored using a custom mySQL database. Upon completion of the data collection a backup copy of the original mySQL database was made and stored on a secure server. The data was exported into Microsoft Excel and SPSS for data analysis.

The presence or absence of the Google.cn filtering statement was used to determine examine if a keyword search or a URL site search was filtered. The statement reads:

据当地法律法规和政策，部分搜索结果未予显示。

[Translation: According to local laws, regulations and policies, some search results are not displayed.]

This dichotomous data – whether the filtering statement was present or not – was analysed using the McNemar test and the Cochran Q test.

In simple terms, the McNemar test can be viewed as a type of chi-square test that uses dependent (i.e., correlated or paired) data rather than independent (unrelated) samples. The McNemar test is a non-parametric statistical test; i.e., it is distribution free and can be used with data sets and samples that are not normally distributed. (Adedokun & Burgess, 2011, p. 126)

The McNemar test, therefore was used to analyse the statistical significance of language on the rate of filtering. The Cochran Q is also a statistical test for dichotomous variables that is related to the McNemar test; however, where the McNemar test is limited to two dichotomous variables, the Cochran Q tests whether the proportions of three or more

dichotomous variables are equal when these variables have been measured across the same cases (Seeger & Gabrielsson, 1968). In this study, the Cochran Q test is used to analyse statistical significance of the rate of filtering across the seven runs.

Keyword search results and site search results were further classified as not filtered, sometimes filtered, or always filtered over the seven runs:

- If a URL search result **contained** the Google filtering statement **every time** it was searched, it was labeled as always filtered.
- If a URL search result **did not contain** the Google filtering statement **any time** it was searched, it was labeled as not filtered.
- If a URL search result was found, **at least once, to contain** the Google filtering statement **and** found, **at least once, not to contain** the Google filtering statement, it was labeled as sometimes filtered.

The descriptive analysis of the data was reported using counts and percentages. Counts and percentages were used to track the pervasiveness of filtering of keyword searches, URL site searches, and overblocking.

## **Chapter 4 Results**

This chapter presents the results of the study in four sections in response to the four research questions relating to pervasiveness and consistency of filtering and the material removed during filtering.

### **4.1 Frequency of Filtering of Search Results**

Research Question 1: How pervasive was the filtering of search results on Google.cn?

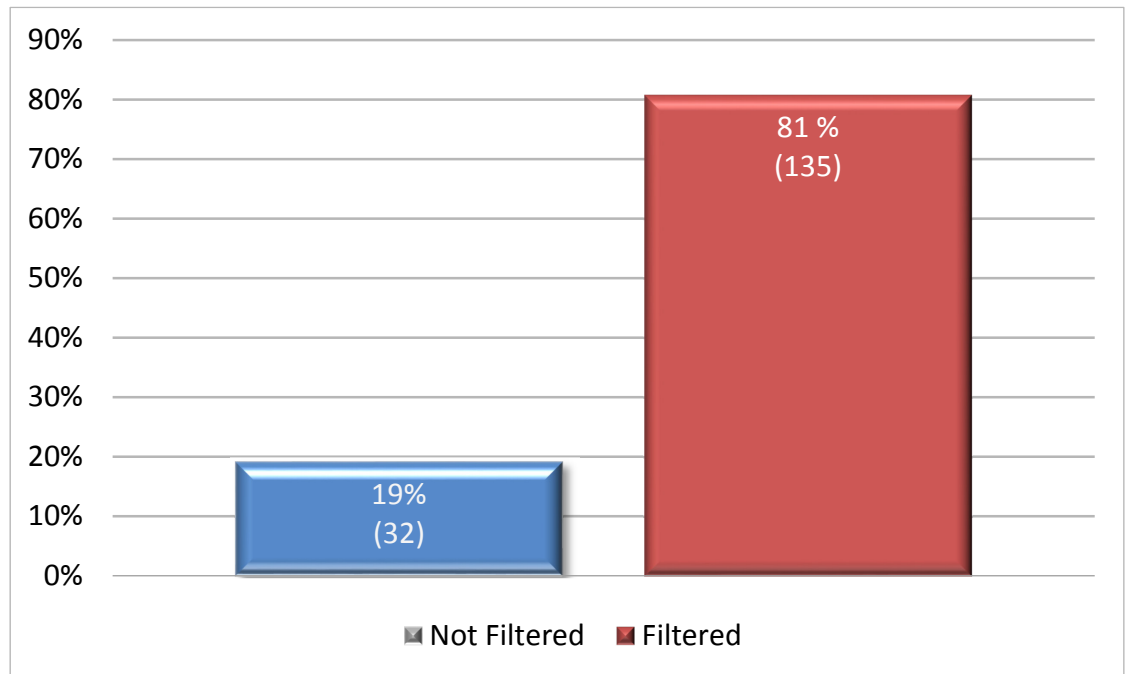
To respond to the first research question on pervasiveness, percentages were used to compare the proportions of keywords always filtered, sometimes filtered, and not filtered. Percentages provide a consistent measure of comparison given that there are unequal numbers of keywords which are grouped by language and/or by the general topics (Falun Gong or Tibet). As stated earlier, the keywords relating to the topics of Falun Gong and Tibet were chosen because these subjects are known to be considered highly sensitive by the Chinese Government. As well, Chinese and English keywords related to each topic were included to examine what impact language might have on search filtering.

All keywords were searched seven times over the course of data collection. Initially 180 keywords were included in the data collection; 13 keywords were removed from this analysis due to errors in the data collection, either human or mechanical. Because data collection was a manual process done by volunteers, there were some errors introduced into the results. For example, volunteers incorrectly pasted the search URL instead of the search result page's HTML source code. Mechanical errors were caused primarily by subtle variations in Google source code and by symbols, particularly apostrophes, in URLs and

irregularities in source code, which caused errors during parsing and analysis by the data tool. The data analysis is based on the remaining 167 keywords.

The overwhelming majority of keywords were filtered at least once; 81% (136 of 167) of the keywords searched in Google.cn returned Google's filtering statement with the search results at least one of the seven times searched (see Figure 6).

Figure 6 Percentage of all keyword searches filtered at least once

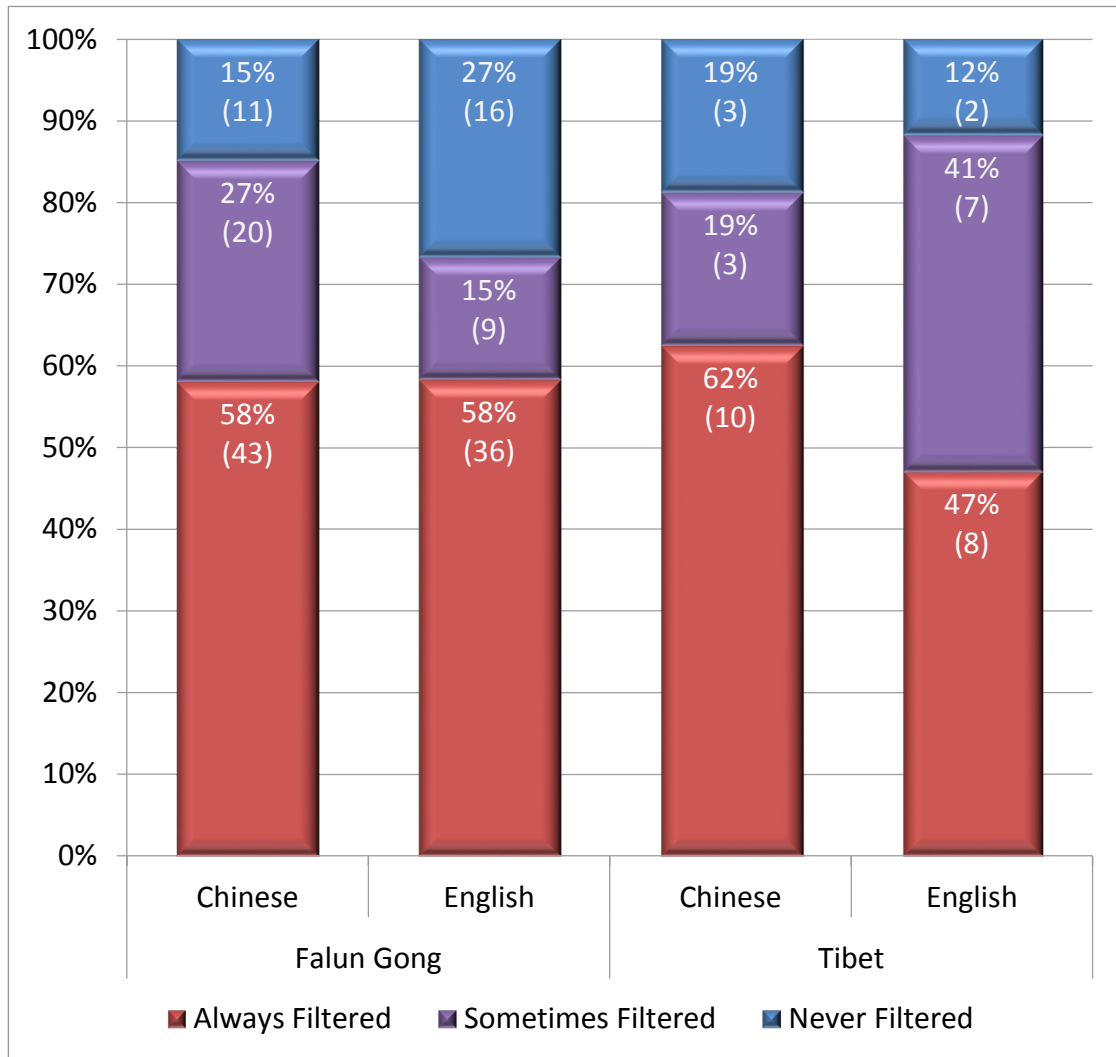


The most filtered keyword subsets by topic and language in order are (see Figure 7):

- English language Tibet keywords: 88% filtered
  - 41% sometimes filtered, 47% always filtered
- Chinese language Falun Gong keyword: 85% filtered
  - 27% sometimes filtered, 58% always filtered
- Chinese language Tibet keyword searches: 81% filtered
  - 19% sometimes filtered, 62% always filtered

- English language Falun Gong keyword: 73% filtered
  - 15% sometimes filtered, 58% always filtered

Figure 7 Frequency of keyword search filtering by language and topic



#### 4.2 Consistency of Filtering

Research Question 2: How consistent was the filtering of search results on Google.cn?

To determine whether each of the runs of keyword searches were equally filtered, two Cochran Q tests were performed. The first examined the filtering of the Chinese

keywords across the seven runs and the second examined the filtering of the English keywords across the seven runs.

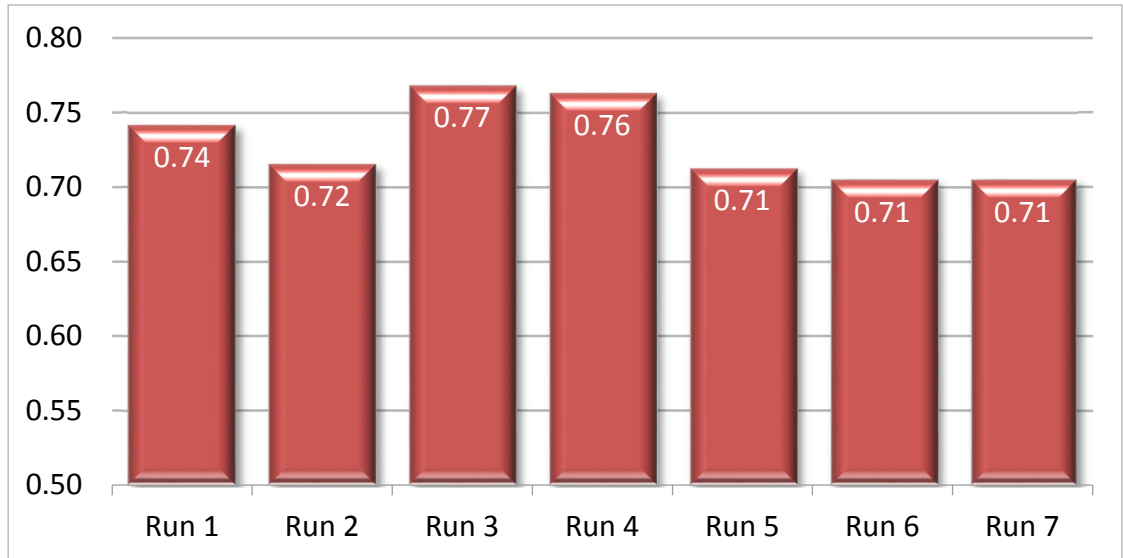
Of the 90 Chinese keywords, there were no missing values across the seven runs. The proportions ranged from 0.71 to 0.77 (see Table 3, Figure 8). The Cochran Q test did not indicate any significant differences among the seven proportions:  $\chi^2(6) = 9.06, p = 0.17$ . In other words, the seven runs completed using Chinese keywords were equally filtered.

Table 3 Cochran Q mean values for Chinese keywords search by run

Run #	<i>n</i> *	Mean	Std. Deviation
Run 1	93	0.74	0.44
Run 2	95	0.72	0.45
Run 3	95	0.77	0.42
Run 4	93	0.76	0.43
Run 5	94	0.71	0.45
Run 6	95	0.71	0.46
Run 7	95	0.71	0.46
Valid N (listwise)	90		

\* *n* = number of keywords

Figure 8 Cochran Q mean values for Chinese keywords search by run



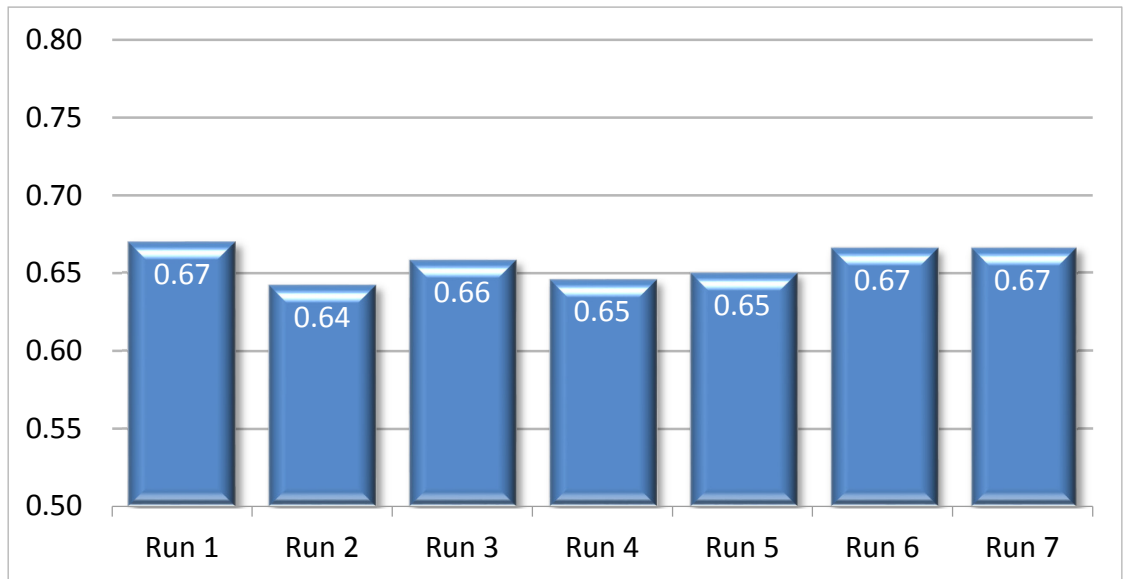
A second Cochran Q Test was performed with 78 English keywords with no missing values across the seven runs. The proportions ranged from 0.64 to 0.67 (see Table 4, Figure 9). The Cochran Q test did not indicate any significant differences among the seven proportions,  $\chi^2(6) = 1.38, p = 0.97$ . Like their Chinese counterparts, the seven runs completed in English were equally filtered.

Table 4 Cochran Q mean values for English keywords search by run

Run #	<i>n</i>	Mean	Std. Deviation
Run 1	85	0.67	0.47
Run 2	84	0.64	0.48
Run 3	85	0.66	0.48
Run 4	82	0.65	0.48
Run 5	83	0.65	0.48
Run 6	84	0.67	0.47
Run 7	84	0.67	0.47
Valid N (listwise)	78		



Figure 9 Cochran Q mean values for English keywords search by run



### 4.3 Frequency of Filtering by Language

Research Question 3: Did language affect the frequency of filtering of searches on Google.cn?

A McNemar test was performed in order to examine whether the language of the keyword (English or Chinese) affected how frequently keywords were filtered on Google.cn over the seven search runs.

After removing keywords with missing values, fifty-nine paired keywords remained to test whether Chinese keywords relating to the sensitive topics of Falun Gong and Tibet in China were more filtered than those same words in English. There were seven iterations of each keyword search and five of the seven iterations (71%) of the exact McNemar tests found there was a significant difference ( $p < 0.05$ ) between the proportion

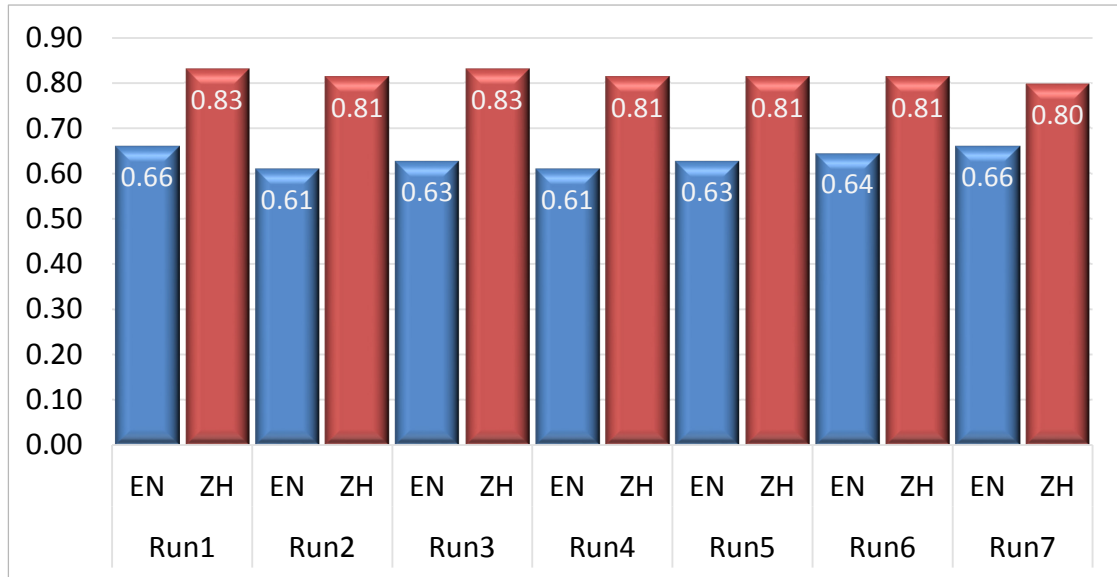
of filtered keywords in English and Chinese. The Chinese words tended to be more frequently filtered than their English counterparts (see Table 5, Figure 10).

Table 5 McNemar mean values for English (EN) and Chinese (ZH) keyword searches by run

Run #	Language	Mean	Std. Deviation	<i>p</i> value
Run 1	EN	0.66	0.48	0.06
	ZH	0.83	0.38	
Run 2	EN	0.61	0.49	0.02 *
	ZH	0.81	0.39	
Run 3	EN	0.63	0.49	0.01 *
	ZH	0.83	0.38	
Run 4	EN	0.61	0.49	0.02 *
	ZH	0.81	0.39	
Run 5	EN	0.63	0.49	0.04 *
	ZH	0.81	0.39	
Run 6	EN	0.64	0.48	0.04 *
	ZH	0.81	0.39	
Run 7	EN	0.66	0.48	0.12
	ZH	0.80	0.40	

Note. \* significant at the  $p < 0.05$  level

Figure 10 Proportion of filtered English (EN) and Chinese (ZH) keywords per run



#### 4.4 What was Filtered?

Research question 4: When filtering occurred, what was filtered from search results on Google.cn?

To determine what organisations were filtered from Google.cn search results, analysis focused on tracking the presence of the Google filtering statement. Whereas the analysis described in the previous sections analysed the filtering of keyword searches through the presence or absence of the Google.cn filtering statement, this analysis examined the site search results for the presence or absence of the Google.cn filtering statement to examine what content was being removed from those keyword searches; where the former sought to ascertain *whether* filtering occurred, this analysis seeks to ascertain *what* is being filtered. The list of URLs in this section was compiled by comparing the top ten search results from Google.cn keyword searches with the top ten results from

Google.ca. URLs found in Google.ca keyword search results but not in Google.cn keyword search results were then searched one-by-one in Google.cn to examine whether or not each URL was filtered in Google.cn. The URLs are only included if they were generated from keyword searches that were filtered in Google.cn.

Comparing the top ten keyword search results from Google.ca to those from Google.cn, there were 7241 instances of missing URLs. Of these 635 (~10%) of the URLs could not be analysed due to human or mechanical error in the data collection, leaving 6606 URLs remaining. Many URLs appeared more than once because the keywords were each run seven times, many keywords were searched in both Chinese and English, and many keywords have similar or related meanings. Duplicate URLs and URLs from unfiltered keyword searches were removed. This subset of 1532 unique URLs from filtered keyword searches is the basis of this analysis.

Instances of overblocking, cases “where sites with superficial similarities to those with sensitive material, but different content, were filtered” or officially sanctioned web pages with content against sensitive topics were filtered, were also uncovered and are presented in subsection 4.4.2 (OpenNet Initiative, 2006, p. 23). Interestingly, cases were identified of individual site search results where the Google filtering statement appeared and the URL was also in the site search results. Though relatively few, they are notable nonetheless. They are presented in subsection 4.4.3.

#### **4.4.1 *Summary of Unique URLs Filtered in Google.cn***

Mirroring the presentation of the results from the OpenNet Initiative (2006), results are divided by always filtered and sometimes filtered URLs.

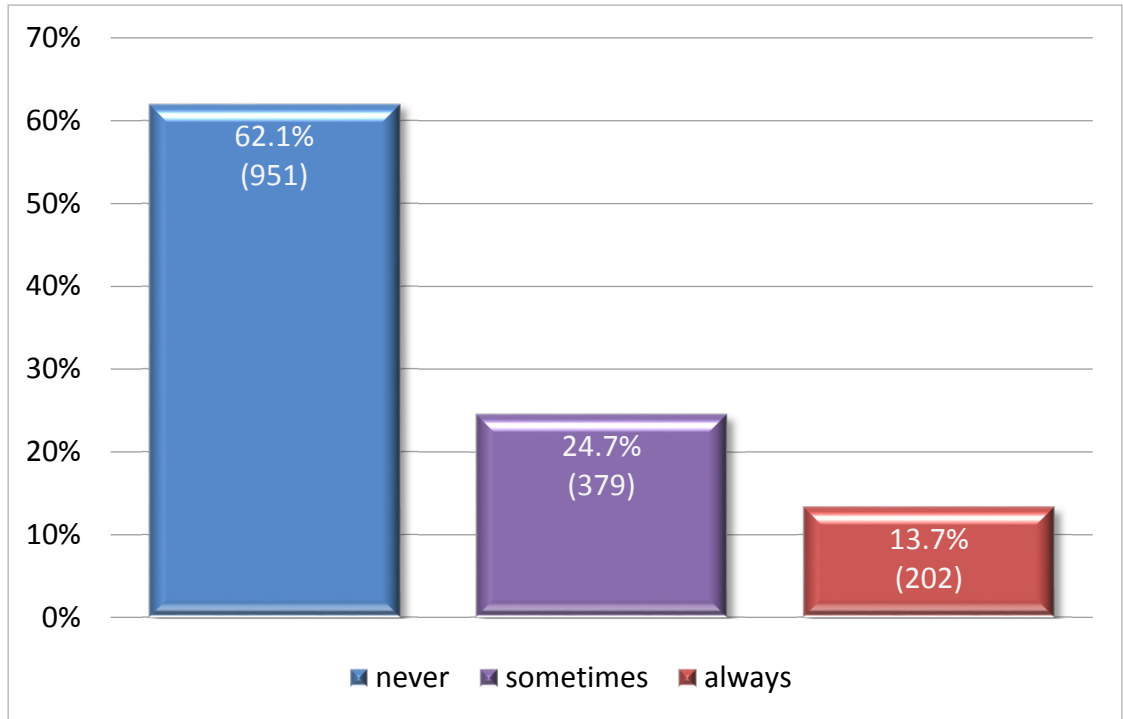
URLs are collected into organisations by related domains. This could be multiple domains for a single organisation differentiated by service or language, or may be alternate domains like those mentioned earlier from the Citizen Lab (2007). For instance, there were 60 unique URLs from 15 separate domains relating to the Epoch Times amongst the URLs from filtered keyword searches. Grouping domains together according to the organisation they represent provides a clearer picture of the affected parties’ tendency to be filtered across many individual URLs and domains. These organisations also allow for concision in the results tables.

Overall, a total of 581 URLs were subject to filtering at least once, 202 unique URLs were always filtered and 379 unique URLs sometimes filtered (see Table 6, Figure 11). The 951 unfiltered URLs are not included in the analysis below.

Table 6 Total number of unique URLs by organization and domain from all filtered keyword searches

<b>Rate of Filtering</b>	<b>Organisations</b>	<b>Domain</b>	<b>Unique URLs</b>
Always	63	86	202
Sometimes	28	100	379
Not filtered	581	692	951
<b>Total</b>	<b>672</b>	<b>878</b>	<b>1532</b>

Figure 11 Percentage of filtered unique URL site searches



Of the unique URLs stemming from Falun Gong keywords, several things can be said. Relating to Falun Gong, 186 unique URLs from 83 domains grouped into 60 organisations were always filtered and, 250 unique URLs from 75 domains grouped into 25 organisations were sometimes filtered (see Table 7). Of these, FalunDafa.org, ClearHarmony.net, Zhengjian.org and Boxun.com were the organisations most often always filtered, while Minghui.org, Epoch Times, Creaders.net and Wikipedia were the organisations most often filtered only sometimes. (See Appendix C Table 13 and Table 14 for more details.)

Table 7 Number of unique URLs by organization and domain from filtered Falun Gong keyword searches

<b>Rate of Filtering</b>	<b>Organisations</b>	<b>Domain</b>	<b>Unique URLs</b>
Always	60	83	186
Sometimes	25	75	250
<b>Total</b>	<b>85</b>	<b>158</b>	<b>436</b>

Originating from Tibet keyword searches were 30 unique URLs, which were from 13 domains representing 13 organisations; these were always filtered. As with the Falun Gong URLs, there were fewer domains and organisations that were only sometimes filtered: 12 unique URLs from five domains grouped from four organisations (see Table 8). Unlike the Falun Gong data, where there were more unique URLs were sometimes filtered, in the Tibet keyword searches more unique URLs were always filtered than sometimes filtered. The organisations most often always filtered were BBC, Central Tibetan Administration, SaveTibet.org and the International Tibet Independence Movement; those most often filtered only some of the time were Epoch Times, Wikipedia, China.com and MySpace (see Appendix C Table 15 and Table 16 for more details).

Table 8 Number unique URLs by organization and domain from filtered Tibet keyword searches

<b>Rate of Filtering</b>	<b>Organisations</b>	<b>Domains</b>	<b>Unique URLs</b>
Always	13	13	30
Sometimes	4	5	12
<b>Total</b>	<b>17</b>	<b>18</b>	<b>42</b>

#### 4.4.2 Overblocking

Overblocking fell into 2 main categories: *unrelated material* and *anti-Falun Gong/Tibet material*. In total 35 of the 581 unique filtered URLs were subject to overblocking, in other words, 6% of the unique URL filtering on Google.cn was overblocking.

Overall, 16 unique URLs from 11 domains were subject by overblocking of *anti-Falun Gong/Tibet material* (see Table 9). Among the most notable examples were CCTV's (the official Chinese's state broadcaster) official TV specials on the Falun Gong and Mingjing.org, a China based web portal with anti-Falun Gong materials (see Appendix C Table 17 for more details).

Table 9 Organisations with anti-Falun Gong/Tibet material affected by overblocking

Organisations	Domains	URLs
202.84.17.11	1	1
bbsland	1	1
CCTV	2	3
China.com	1	1
Creaders.net	1	1
Exposing the Falun Gong	1	1
Mingjing.org	2	6
Network54.com	1	1
Sina.com	1	1
<b>Grand Total</b>	<b>11</b>	<b>16</b>

A number of unique URLs ( $n = 19$ ) from 13 domains which contained material unrelated to Falun Gong or Tibet were filtered (see Table 10). Overblocked sites included the municipal webpage of Falun, Sweden, which shares part of Falun Gong's name, and HongZhi, a Stone manufacturer and exporter in China, whose company name is the same



as the family name of the founder of Falun Gong (see Appendix C Table 18 for more details).

Table 10 Organisations unrelated materials affected by overblocking

<b>Organisation</b>	<b># Domains</b>	<b># URLs</b>
BBC	1	2
Britannica	1	1
China.com	1	1
ChinaDaily.com	1	1
Columbia University	1	1
DAFA (Direct Aid for Africa)	1	1
Falun Municipality, Sweden	1	2
flowingdata.com	1	1
National Solar Observatory	1	1
phonecard.dajiyuan.com	1	3
Qianjinfang.net	1	1
tradekey	1	1
Wikipedia	1	3
<b>Grand Total</b>	<b>13</b>	<b>19</b>

#### **4.4.2.1 URLs that were Simultaneously Filtered and Available in Search Results**

Unexpectedly, there were instances ( $n = 61$ ,  $< 1\%$ ) where a site search returned both the URL searched in the results and Google’s filtering statement. In other words, in a very small number of instances, the search, though filtered for the result for which it searched, also concurrently returned that very search result. This appears inherently contradictory, as even though the Google.cn search results stated the search had been filtered, the URL searched was available in the search results.

It should be noted that there is no division by category in the following because all of these results relate to Falun Gong keywords. No URLs relating to Tibet keywords were

simultaneously filtered and available in the list of results when searched. Of the URLs that were both available in the search results and returned the Google.cn filtering statement, 47.4% were overblocked (unrelated to Falun Gong), 21.1% were anti-Falun Gong sites and 31.6% were Falun Gong sites (see Table 11, see Appendix C Table 19 and Table 20 for more details).

Table 11 Number and percentages of unique URLs that were simultaneously filtered and available in site search results

<b>Material type</b>	<b># Unique URLs</b>	<b>% Unique URLs</b>
Anti-Falun Gong	4	21.1%
Overblocked	9	47.4%
Falun Gong	6	31.6%
<b>Total</b>	<b>19</b>	<b>100.0%</b>

## Chapter 5 Discussion

Research on Internet filtering in China has often used Google as a tool to measure Chinese censorship practices. More recently, as companies running search engines have begun to comply with the PRC's regulations, focus has turned to the internal filtering practices of individual search engines (see Herold and de Sela, 2015; Amnesty International, 2006; The Internet in China, 2006; Human Rights Watch, 2006; Villeneuve, 2008).

This study examined Google's filtering practices from an end user perspective by tracking the presence of the Google.cn filtering statement. The filtering of Falun Gong and Tibet keyword searches on Google.cn was found to be both pervasive and consistent. Over 80% of the keyword searches were subject to filtering. The findings of this research support previous studies which examined the PRC own internet censorship practices to filter wide amount of materials.

Gauging the impact on the content of search results, i.e., what or how much is being filtered, is more difficult to determine. Where earlier studies compare Google.com search results inside and outside of the Great Fire Wall of China to examine what was being censored (OpenNet Initiative, 2006; Zittrain & Elderman, 2003), this study compared Google.ca top ten search results and Google.cn top ten search results to examine what Google.cn was filtering. Filtering affected 38% of the unique URLs compiled by comparing these top ten searches.

This chapter first discusses the findings from the research questions followed by research design issues, contributions and an outline of future research directions.

## 5.1 Research Questions

The pervasiveness and consistency of China's own filtering practices vary due to the multitude of methods employed in filtering (Crandall et al., 2007; Faris & Villeneuve, 2008; OpenNet Initiative, 2006) and the many different regulatory bodies, statutes and policies governing the Internet (Hughes & Wacker, 2003; OpenNet Initiative, 2006; Wacker, 2003). Beginning with Google Inc.'s public announcement in January 2006, Google did its own internal filtering of Google.cn in accordance with the local laws of the PRC.

### 5.1.1 *Pervasiveness of Filtering*

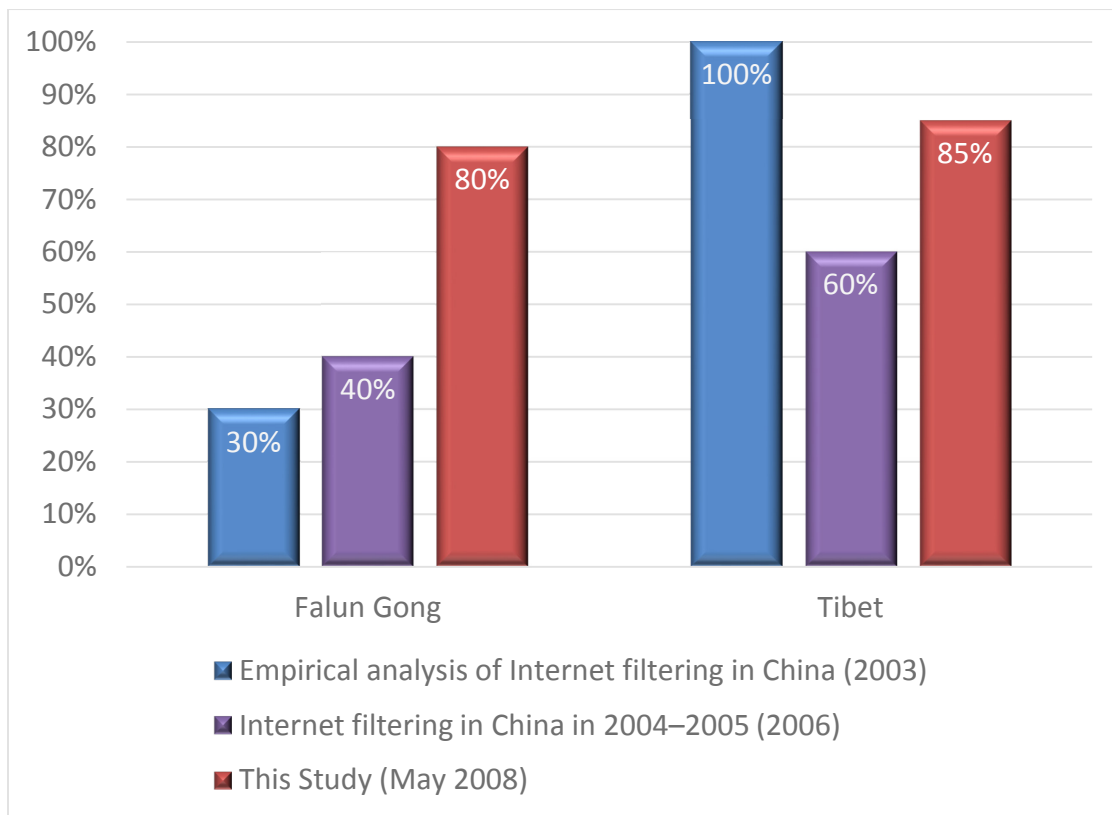
Research Question 1: How pervasive was the filtering of search results on Google.cn?

This study confirmed that Google.cn was pervasive in filtering the sensitive topics Falun Gong and Tibet; 80% of the Falun Gong keywords and 85% of the Tibet keywords were affected by filtering. As part of *IFinC* (OpenNet Initiative, 2006) the authors presented a longitudinal table comparing their results with the earlier *EAIFinC* (Zittrain & Elderman, 2003). The longitudinal analysis was of "Proportion of Top 10 and Top 100 Google Search Results Inaccessible" by topic, including Falun Gong and Tibet. The filtering rates related to Falun Gong and Tibet have been extracted and compared to the overall filtering rates found in this study. Though a slightly different metric, since these earlier studies also track filtering on a Google top ten search, it does give an overall way to compare filtering rates by PRC imposed on search results by rendering the URLs inaccessible to the rate of filtering done by Google Inc. in house on Google.cn. Figure 12 uses the proportion of top ten Google Search Results Inaccessible for Falun Gong and Tibet to compare with the findings in this study.

The rate of filtering with respect to Falun Gong increased dramatically from the two earlier studies. The filtering of Falun Gong materials had only increased, first a 10% increase between *EAI FinC* to *IFinC*, then a doubling from 40% to 80% between *IFinC* and this study. Initially material related to Tibet had been 100% filtered according to the *EAI FinC* and dropped to 60% filtered in the *IFinC* study. This study saw another increase in the filtering rate to 85%.

The fact that in the *EAI FC* Tibet was filtered 100% shows that PRC had the technical ability to block access completely, and was willing to do so without consideration for nuances in content. Not all keyword content is sensitive. Material would also be related to other overlapping topics, such as the Tibetan Autonomous Region, tourism, history, anti-independence materials, etc. As the *IFinC* authors concluded, “this likely indicates China’s willingness to tolerate blocking unrelated content to prevent access to sensitive materials” (OpenNet Initiative, 2006, p. 23). The increase in Falun Gong filtering may indicate greater attention being paid by authorities to this topic as time went on, which is consistent with Tong (2009) conclusions about the rising persecution of Falun Gong in China. The rate of filtering of about 80 to 85 percent for sensitive topics found in this study are in line with Ang’s arguments that “the censor’s goal is achieved if it is difficult for *most* users *most* of the time to access *most* of the material” (2005, p. 40). This is represented by the compromise made by Google. PRC is satisfied as long as the material is difficult to access and is willing to tolerate overblocking. Google is also interested in the access to information as it is a fundamental part of the product they are offering the end user. Unlike the PRC, Google is invested in filtering only sensitive content, notifying the user that filtering has occurred and maintaining access to the non-sensitive materials.

Figure 12 Comparison of Google.cn keyword search filtering rates to proportion of top ten Google search results inaccessible in previous studies



### 5.1.2 Consistency of Filtering

Research Questions 2: How consistent was the filtering of search results on Google.cn?

This study found that filtering practices on Google.cn were consistent. Over the seven search run there was no significant difference in the filtering of keyword searches. The consistency of filtering was examined by the language of the keyword. The two subsets of Chinese keywords and English keywords were nearly equal as 52% of the keywords are Chinese and 48% are English. Throughout *IFinC* (OpenNet Initiative, 2006) the consistency of filtering by the PRC varies by method and topic; generally Falun Gong and Tibet appear to be consistently filtered by the various methods tested in the study. Google.cn’s filtering of these topics is certainly more consistent overall than the mix of

methods used by the PRC. However, given that this study limited its investigation to two topics known to be highly sensitive and heavily censored, it is difficult to say if the consistency of Google.cn's filtering is due to Google's own filtering practices or if the sensitive character of the topics means that Google was given a near exhaustive keyword list to filter by the PRC.

### **5.1.3 Language and Filtering**

Research Question 3: Did language affect the frequency of filtering of searches on Google.cn?

While the overall filtering of keyword searches by language was consistent over the seven runs, the pervasiveness of filtering was not. When the filtering rates of paired Chinese keyword and English keyword searches were compared, the Chinese keyword searches were significantly more filtered in five out of the seven search runs. This difference in the keyword search filtering rate bears out the literature from the Citizen Lab (2007), Weiner (2011) and Qiang (2011), which all point to using an alternate language or wordplay to subvert censors.

Again, *IFinC* tested a slightly different metric, the accessibility of Google top ten search results. While they did keyword searches in both Chinese and English, their measures were based on whether the language of the inaccessible URL was Chinese or English. Generally in their results, Chinese language sites were found to be less accessible than English language sites. The grouping for categories was slightly different in this section of *IFinC*, Independence movements were a category, which included Tibet amongst others:

A relatively small percentage of Chinese language sites listed for searches for “Tibet” and “Taiwan” were inaccessible; the number rose significantly when the search query included the term “independence.” Chinese language sites related to Tibet were slightly less accessible than sites discussing Taiwan; this gap was more pronounced for English language search results. (OpenNet Initiative, 2006, p. 33)

Falun Gong was its own category and “[unlike] in many other categories of content, a significant percentage of the English language sites were blocked, in both absolute terms and relative to Chinese language sites” (OpenNet Initiative, 2006, p. 29).

#### **5.1.4 Content Removed and Overblocked**

Research Question: When filtering occurred, what was filtered from search results on Google.cn?

This study found that many of the URLs that were filtered were related to the same domain and many of the domains were alternate domains all related to the same organization. This finding held for URLs that were always filtered and sometimes filtered. This finding was unsurprising as using alternate domains is a well-established technique for circumventing Internet censors (Citizen Lab, 2007).

The organisations identified as the most often always filtered or sometimes filtered are large well known and well established web sites with a diverse range of content. The most affected sites were either web portals related to either Falun Gong or Tibet (FalunDafa.org, ClearHarmony.net, Zhengjian.org, Central Tibetan Administration, Minghui.org, SaveTibet.org and the International Tibet Independence Movement) or news sites or portals (Epoch Times, Boxun.com, BBC, Creaders.net, China.com), and of course,



Wikipedia which was only sometime filtered for both Falun Gong and Tibet related searches.

The overall rate of overblocking found in this study was only 6.0% on Google.cn. This is much lower than previous studies have found, notably the overblocking testing in *IFinC* and by the ConceptDoppler team (OpenNet Initiative, 2006, pp. 42-45; Crandall et al., 2007).

What is less clear is why Google.cn was much more precisely filtered than the PRC techniques achieved. This may have more to do with Google Inc. interests in Internet filtering vs. PRC interests in Internet filtering than any technical capability. The PRC “employs a sophisticated infrastructure that filters content at multiple levels and that tolerates overblocking as the price of preventing access to prohibited sites” (OpenNet Initiative, 2006, p. 52). For Google, overblocking undermines the product they are offering as a company: search results to users and detailed profiles of users to advertisers based on detailed and accurate tracking of user interests and preferences. In documents about the decision to filter Google.cn, Google repeatedly points to transparency as a mitigating factor in its decision and a ground-breaking step: “Google.cn presents to users a clear notification whenever links have been removed from our search results in response to local laws and regulations in China. We view this a step toward greater transparency that no other company has done before” (Schrage, February 15, 2006). Google contrasts its methods to the methods of other companies and to the methods of the PRC:

Importantly, China’s filtering efforts lack transparency: the state does not generally admit to censoring Internet content, and consequently there is no list of banned sites and no ability for citizens to request reconsideration of

blocking, as some other states that filter provide. The topics defined as sensitive, or prohibited, by China's legal code are broad and non-specific, and enforcement of laws such as the ban on spreading state secrets discourages citizens from testing the boundaries of these areas. China's legal and technological systems combine to form a broad, potent, and effective means of controlling the information that Chinese users can see and share on the Internet. (OpenNet Initiative, 2006, p. 52)

This control is precisely the Chinese government's motivation. Though Google is engaging in online censorship, the company is motivated by corporate interests and the perpetuation of the company, whereas the PRC is motivated by the perpetuation of the state.

## **5.2 Research Design Issues**

This study used 180 keywords limited to two topics known to be considered sensitive by the PRC. This somewhat limits the ability to generalise the results. A larger sample would have allowed for more specific analyses. Likewise a longitudinal study preformed over two different periods would have provided more objective data. Due to resource limitations this was not feasible. Also due to resource limitations, most research assistants were drawn from the local Halifax area.

There are a number of difficulties with any study involving Google, as the inner workings of Google's search engine are corporate trade secrets and thus not known to researchers. This study assumes that Google.cn is censored the same way by Google Inc., regardless of the location of the searcher. In addition, the 'completeness' of any set of results is impossible to know since the number of servers searched, how much of any

database is contained on any given server, when the database is updated, and how often individual servers are updated are all unknown. Of the little information available the *Search Monitor Project* does note that “[some] servers for google.cn are hosted inside China, but some google.cn servers are located outside China and can be queried from outside China” (Villeneuve, 2008, p. 9).

As noted in the introduction there are a number of a limitations when trying to draw clear conclusions about why a URL appears in Google.ca results and not Google.cn. First, a given URL may have not ranked high enough to appear in the top ten results but may have been included in the overall results list. Second, the URL may not have been index in the Google.cn database, or may not have been included on the servers searched. In order to verify that a URL was actually filtered, this study searched these absent URLs again in Google.cn using a site search for the specific URL. While care was search each missing URL using a site search operator, it is acknowledged that the results of this study are suggestive.

### **5.3 Contributions**

The current research makes four contributions to research relating to Google’s Internet filtering practices on Google.cn.

First and foremost, this study provided analysis on Google Internet filtering activities on Google.cn. Given the short time Google.cn was actually online as a filtered search engine, there was only a limited window for research on Google’s internal filtering practices to be done. The closest study to this study is the *Search Monitor Project* (Villeneuve, 2008). That study was also collecting data about the same period, however there was follow-up work from both this study and from Villeneuve’s that can no longer

be done as Google no longer maintains a Google.cn as a filtered search engine in accordance with PRC regulations.

Second, this study follows up on the earlier work on Internet filtering in China (OpenNet Initiative, 2006; Zittrain & Elderman, 2003). Adapting their methodology using Google.com as a tool to measure Chinese Internet filtering, this study examines Google's own internal filtering practices under Chinese law. This study was able to determine that while Google's rates overblocking was lower than PRC, the overall filtering rates were in line with those found in the ONI's earlier studies.

Third, this study contributes to the research on Falun Gong and Tibet. The persecution of Falun Gong and Tibetan Independence groups can vary based on local and world events. Likewise determining what the rate of Internet filtering is by the Chinese government or their partners, like Google Inc., can be quite difficult. This study provides specific data about the Internet filtering of the topics Falun Gong and Tibet in May 2008, immediately prior to the 2008 Summer Olympics in Beijing.

Fourth, this study is unique in analysing of overblocking on Google.cn. Overblocking The *Search Monitor Project* (Villeneuve, 2008) raised the possibility of overblocking, but did not investigate it. Despite Google sophisticated technology, 6% of material filtered was either unrelated or material anti to Falun Gong. Not only is internet filtering difficult to track it is difficult to execute with complete precision. Internet filtering impact extends beyond the target, harmful to access to information not intended to be filtered. As discussed while PRC is comfortable blocking more information to rather than let any sensitive information through, this is directly counter to Google's stated missions of satisfying users and providing access to information (Schrage, 2006).

## 5.4 Future Research

Any new study conducted would have to be modified because Google.cn no longer exists as a filtered regional Google database. The Google.cn page now redirects visitors to Google.com.hk, Google's regional search engine for Hong Kong. There are a number of avenues of investigation that could be pursued with the original data collected.

Though beyond the scope of this thesis, due to the richness of the data collected, more analysis to respond to further research questions could be conducted.

- Did the ordering of the URLs within the top ten results affected rates of filtering?
- Did the source of the keyword affect the frequency of filtering of searches on Google.cn?
- Did the number of times a URL appeared correlate to the rate of filtering on Google.cn?
- Did English language keyword search return as significantly higher number of search results than its Chinese translation?
- Did Google.ca keyword search return as significantly higher number of search results than the same keyword search on Google.cn?

## Chapter 6 Conclusion

It is clear from this study that Google Inc. was true to its goal of “greater transparency” as the filtering statement was clearly visible and easy to track (Schrage, 2006). However, while Google was true to its mission of increased transparency, its filtered search engine actually hid more information on Falun Gong and Tibet than previous ONI studies found China did. Earlier research on China’s Internet censorship clearly establishes that the practice does not have to be complete or infallible as “the censor’s goal is achieved if it is difficult for *most* users *most* of the time to access *most* of the material” (Ang, 2005, p. 40). Fulfilling China’s censorship goal is inherent in the compromise made by Google in launching Google.cn. Google is also interested in its own access to user search information as a fundamental part of the product they are offering the end user. PRC is satisfied as long as the material is difficult to access and is willing to tolerate overblocking. Unlike the PRC, Google is invested in filtering only sensitive content, notifying the user that filtering has occurred and maintaining access to the non-sensitive materials. This is in line with the results of the earlier ONI studies on censorship of the Internet in China and the culture of censorship created by the PRC, that filtering need not be complete or infallible to be effective.

However, as Villeneuve points out, this does not account for the fact Google was the main source of information on the sensitive topics. No Chinese search engine was able to provide the information Google.com provided. Google’s popularity as a search engine rested on the fact that its results were unfiltered. Even after the launch of Google.cn, Google.com was preferred over the new Chinese regional search engine (Qiu, 2010). Google.cn filtering was pervasive and consistent, and it was significantly higher for

Chinese language keywords than for English language keywords. For these reasons, Google.cn was not providing any more information than any other Chinese based search engine to Chinese citizens seeking information. Google's filtering statement indicated to searchers that results were being filtered, but it gave no indication of what or how much was missing. The filtering statement could be seen by citizens as a warning that they were searching a topic that might put them at risk.

It is impossible to know from this study if the filtering statement acted as a deterrent to further searches on sensitive topics or encouraged searchers to try alternate methods to try to gain access to information that was in the search engine database but was being kept from them. Many citizens would be aware that searching sensitive topics would put them at risk and that repeated searching could be tracked. So Google's action could be seen as a service to searchers—the transparency or openness about sites that existed but could not be revealed in the search engine. Or Google's action could be seen as participating in the creation of a chilling effect—warning searchers not to try to find information, when that activity might get them into trouble. As Villeneuve points out,

Although the total number of censored sites is not high, especially when compared to the amount of indexed sites, the significance of these sites in providing alternative information should not be underestimated. These censored sites are often the only sources of alternative information available in the top ten results for politically sensitive search queries. ... Although, these search engines censor less content than the domestic Chinese search engine Baidu, the removal of these sites from the search engines has an

unambiguous, negative impact on the freedom of expression. (Villeneuve, 2008, p. 3)

Google Inc.'s filtering of Google.cn was found to be pervasive and consistent. However, this complicated relationship with the People's Republic of China ultimately proved unsustainable, evident in Google Inc.'s choice to shut down its filtered search engine in 2010.



## Appendix A      Keywords

Table 12 List of keywords with translation and topic

ID	PairID/note	Keyword	Translation	Topic
1	Pair01	Be in a daze	发愣	Falun Gong
2	Pair02	Be raised to the skies	升天	Falun Gong
3	Pair03	Buddha stretches a thousand hands	佛展千手法	Falun Gong
4	Pair04	Burn oneself OR self-immolation	自焚	Falun Gong
5	Pair05	Celestial burial	天葬	Tibet
6	Pair06	Celestial wrath	天怒	Tibet
7	Pair07	Choekyi Nyima	确吉尼玛	Tibet
8	Pair08	Clearwisdom or Minghui	明慧	Falun Gong
9	Pair09	Dafa	大法	Falun Gong
10	dup trans	Dafa, or “Great Law”	大法	Falun Gong
11	No pair	Dajiyuan	Dajiyuan	Falun Gong
12	Pair10	Dajiyuan news network or Epoch Times	大纪元新闻网	Falun Gong
13	Pair11	Dajiyuan or Epoch Times	大纪元	Falun Gong
14	Pair59	Dalai	达赖	Tibet
15	Pair12	Dalai Lama	达赖喇嘛	Tibet
16	Pair13	Deliver law	讲法	Falun Gong
17	Pair14	Disciple	弟子	Falun Gong
18	No pair	Epoch Times	Epoch Times	Falun Gong
19	No pair	Epoch Times news Web site	Epoch Times news Web site	Falun Gong
20	Pair15	Erdini Qoigyijabu	额尔德尼·确吉杰布	Tibet
21	Pair16	Exile OR go Into Exile	流亡	Tibet
22	Pair17	Fa	法	Falun Gong
23	Pair60	fa+lun+gong	法十轮十功	Falun Gong
24	Pair18	Faleng	法愣	Falun Gong
25	No pair	Falu	Falu	Falun Gong
26	Pair19	Falun	Falun	Falun Gong
27	Pair20	Falun Dafa	Falun Dafa	Falun Gong

28	Pair21	Falun Gong	falun gong	Falun Gong
29	Pair61	Dharma chakra	法轮	Falun Gong
30	No pair	Falundafa	法轮大法	Falun Gong
31	Pair22	Fazhe	法谿	Falun Gong
32	Pair23	Fazhegong	法谿功	Falun Gong
33	Pair24	Fazhengnian	发正念	Falun Gong
34	dup trans	Flg	法轮功	Falun Gong
35	No pair	Fozhan Qianshou Fa	Fozhan Qianshou Fa	Falun Gong
36	No pair	Gao Jingyun	Gao Jingyun	Falun Gong
37	Pair25	Gedhun Choekyi	根敦·确吉	Tibet
38	Pair26	Gedhun Choekyi Nyima	根敦·确吉尼玛	Tibet
39	Pair27	Gong	公	Falun Gong
41	No pair	Guantong Liangji Fa	Guantong Liangji Fa	Falun Gong
42	Pair28	Guo Ping	郭平	Falun Gong
43	Pair29	Hongzhi	宏志	Falun Gong
44	No pair	Hu Yuehua	Hu Yuehua	Falun Gong
45	Pair30	Law of gong	功法	Falun Gong
46	Pair62	Law of Perforating the Two Poles	贯通两极法	Falun Gong
47	Pair31	Li Hongzhi	李宏志	Falun Gong
48	Pair32	Li Red Piles	李红痔	Falun Gong
49	Pair33	Li Wenbin	李文斌	Falun Gong
50	No pair	Li Zhiying	Li Zhiying	Falun Gong
51	dup trans	liHongzhi	李宏志	Falun Gong
52	Pair63	Lin Shenli	林慎立	Falun Gong
53	Pair34	Lun	伦	Falun Gong
54	Pair35	Lunda	轮大	Falun Gong
55	Pair36	Lunging	轮功	Falun Gong
56	Pair37	Master	师父	Falun Gong
57	Pair38	Master Li	李大师	Falun Gong
58	Pair39	Minghui	明慧网	Falun Gong
59	Pair40	Minghuinews	明慧新闻	Falun Gong
60	No pair	New Tynasty TV Station	New Tynasty TV Station	Falun Gong
61	Pair41	Nine Commentaries	Nine Commentaries	Falun Gong
62	Pair42	Nine Commentaries on the Communist Party	九评共产党	Falun Gong

63	Pair43	Panchen Lama	班禅喇嘛	Tibet
64	Pair44	Practice gong	炼功	Falun Gong
65	Pair45	Preaching the Law	宏法	Falun Gong
66	Pair46	Right argument network	正见网	Falun Gong
67	No pair	Shenzhou Jiachifa	Shenzhou Jiachifa	Falun Gong
69	No pair	Spring Festival Gala of the World's Chinese	Spring Festival Gala of the World's Chinese	Falun Gong
70	Pair47	Teacher Li	李老师	Falun Gong
71	Pair48	Theurgy adding and hold	神通加持法	Falun Gong
72	Pair64	Tibet	Tibet	Tibet
73	Pair49	Tibet china	Tibet china	Tibet
74	No pair	Tibet independence	Tibet independence	Tibet
75	Pair50	Tibet Talk	西藏论坛	Tibet
76	dup trans	Tibetalk	tibetalk	Tibet
77	Pair51	Tibetan independence	藏独	Tibet
78	Pair52	tolerant or tolerate	忍	Falun Gong
79	dup trans	Truth, Compassion, Tolerance	Truth, Compassion, Tolerance	Falun Gong
80	Pair53	Truthfulness, Compassion, Forbearance	真善忍	Falun Gong
81	Pair54	Ustibet	美国西藏	Tibet
82	No pair	Wang Lixiong	Wang Lixiong	Tibet
83	Pair55	Wenyi Wang	王文怡	Falun Gong
84	Pair56	Zhengjian	政见	Falun Gong
85	Pair57	Zhengjianwang	政见网	Falun Gong
86	No pair	Zhenshanren	Zhenshanren (真善忍)	Falun Gong
87	Pair58	Zhuanfalun	转法轮	Falun Gong
88	Pair41	九评	Nine Commentaries	Falun Gong
89	Pair42	九评共产党	Nine Commentaries on the Communist Party, The	Falun Gong
90	dup trans	伦	lun	Falun Gong
91	dup trans	伦公	lungong	Falun Gong
92	dup trans	伦功	lungong	Falun Gong
93	dup trans	伦攻	lungong	Falun Gong

94	Pair03	佛展千手法	Buddha stretches a thousand hands	Falun Gong
95	dup trans	公	Gong	Falun Gong
96	Pair27	功	Gong	Falun Gong
97	Pair30	功法	Law of gong	Falun Gong
98	Pair02	升天	Be raised to the skies or rise To Heaven	Falun Gong
99	dup trans	发伦	Falun	Falun Gong
100	Pair01	发愣	Be in a daze	Falun Gong
101	Pair24	发正念	Fazhengnian	Falun Gong
102	dup trans	发沦	Falun	Falun Gong
103	dup trans	发论	Falun	Falun Gong
104	dup trans	发轮	Falun	Falun Gong
105	Pair09	大法	Dafa, or "Great Law"	Falun Gong
106	Pair11	大纪元	Dajiyuan or Epoch Times	Falun Gong
107	Pair10	大纪元新闻网	Dajiyuan news network or Epoch Times	Falun Gong
108	dup trans	大纪园	Dajiyuan or Epoch Times	Falun Gong
109	Pair06	天怒	Celestial wrath	Tibet
110	Pair05	天葬	Celestial burial	Tibet
111	Pair29	宏志	Hongzhi	Falun Gong
112	Pair45	宏法	Preaching the Law	Falun Gong
113	Pair37	师父	Master	Falun Gong
114	Pair14	弟子	Disciple	Falun Gong
115	Pair52	忍	Tolerant or tolerate	Falun Gong
116	dup trans	抡功	Lungong	Falun Gong
117	dup trans	攻	Gong	Falun Gong
118	Pair56	政见	Zhengjian	Falun Gong
119	Pair57	政见网	Zhengjianwang	Falun Gong
120	Pair08	明慧	Clearwisdom or minghui	Falun Gong
121	Pair40	明慧新闻	Minghuinews	Falun Gong
122	Pair39	明慧网	Minghui	Falun Gong
123	Pair38	李大师	Master li	Falun Gong
124	Pair31	李宏志	Li Hongzhi	Falun Gong
125	Pair33	李文斌	Li Wenbin	Falun Gong

126	Pair32	李红痔	Li Red Piles(Li Hongzhi)	Falun Gong
127	Pair47	李老师	Teacher Li	Falun Gong
128	Pair63	林慎立	Lin shenli	Falun Gong
129	Pair25	根敦·确吉	Gedhun Choekyi	Tibet
130	Pair26	根敦·确吉尼玛	Gedhun Choekyi Nyima	Tibet
131	No pair	歌功颂德	Sing the Praises of Somebody	Falun Gong
132	Pair46	正见网	Right argument network	Falun Gong
133	dup trans	沦	Lun	Falun Gong
134	dup trans	沦公	Lungong	Falun Gong
135	dup trans	沦功	Lungong	Falun Gong
136	dup trans	沦攻	Lungong	Falun Gong
137	Pair17	法	Fa	Falun Gong
138	dup trans	法仑	Falun	Falun Gong
139	Pair19	法伦	Falun	Falun Gong
140	No pair	法十轮	Fa+lun	Falun Gong
141	Pair60	法十轮十功	fa+lun+gong	Falun Gong
142	Pair18	法愣	Faleng	Falun Gong
143	dup trans	法抡	Falun	Falun Gong
144	dup trans	法沦	Falun	Falun Gong
145	dup trans	法论	Falun	Falun Gong
146	Pair22	法滴	Fazhe	Falun Gong
147	Pair23	法滴功	Fazhegong	Falun Gong
148	Pair61	法轮	Falun, or “dharma chakra”	Falun Gong
149	dup trans	法轮功	Flg	Falun Gong
150	Pair21	法轮功	Falun gong	Falun Gong
151	Pair20	法轮大法	Falun dafa	Falun Gong
152	Pair16	流亡	Exile OR go Into Exile	Tibet
153	Pair44	炼功	Practice gong	Falun Gong
154	Pair55	王文怡	Wenyi Wang	Falun Gong
155	Pair43	班禅喇嘛	Panchen Lama	Tibet

156	Pair53	真善忍	Truthfulness, Compassion, Forbearance	Falun Gong
157	Pair07	确吉尼玛	Choekyi Nyima	Tibet
158	Pair48	神通加持法	Theurgy adding and hold	Falun Gong
159	Pair54	美国西藏	Ustibet	Tibet
160	Pair04	自焚	Burn oneself OR self- immolation	Falun Gong
161	dup trans	藏	Tibet	Tibet
162	dup trans	藏独	Tibetan independence	Tibet
163	Pair64	西藏	Tibet	Tibet
164	Pair49	西藏 中国	Tibet china	Tibet
165	Pair51	西藏独立	Tibetan independence	Tibet
166	Pair50	西藏论坛	Tibet talk	Tibet
167	Pair13	讲法	Deliver law	Falun Gong
168	dup trans	论	Lun	Falun Gong
169	dup trans	论公	Lungong	Falun Gong
170	dup trans	论功	Lungong	Falun Gong
171	dup trans	论攻	Lungong	Falun Gong
172	Pair62	贯通两极法	Law of perforating the two poles	Falun Gong
173	Pair58	转法轮	Zhuanfalun	Falun Gong
174	Pair34	轮	Lun	Falun Gong
175	dup trans	轮公	Lungong	Falun Gong
176	Pair36	轮功	Lungong	Falun Gong
177	Pair35	轮大	Lunda	Falun Gong
178	dup trans	轮攻	Lungong	Falun Gong
179	Pair59	达赖	Dalai	Tibet
180	Pair12	达赖喇嘛	Dalai Lama	Tibet
181	Pair28	郭平	Guo Ping	Falun Gong
182	Pair15	额尔德尼·确吉杰布	Erdini Qoigyijabu	Tibet

## Appendix B Data Collection Screen Shots

Figure 13 Google.cn screen capture: Filtered site search, results removed



Figure 14 Google.cn screen capture: Filtered site search, no sites indexed



Figures 12 and 13 show two different outcomes for the search site:http://www.clearwisdom.net. In both cases the filtering statement appeared but in only

one case the site is index. The first indicates that there is a result removed and in the second there is no result to remove but the search is still known to be sensitive under local laws.

Figure 15 Google.cn screen capture: Unfiltered site search, result returned





Figure 16 Google.cn screen capture: Unfiltered Google.cn site search, no sites indexed



## Appendix C Results Tables

Table 13 Count of always filtered Falun Gong URL site search results grouped by organisation

Organisation	Domains	URLs
202.84.17.11	1	1
9-joy.com	1	1
Aboluowang	1	1
Ascension Gateway	1	1
bbsland	1	1
Boxun.com	3	9
CESNUR	1	1
Chapters Indigo	1	1
China News Digest	1	1
ChinaDaily.com	1	2
Chinese Human Rights Defenders	1	1
ClearHarmony.net	5	20
Club 6Park	1	1
Columbia University	1	1
DAFA (Direct Aid for Africa)	1	1
DWNews.com	1	3
everything2.com	1	1
Exposing the Falun Gong	1	1
Falun Dafa in Nevada	1	2
Falun Dafa Information Center	1	1
Falun Dafa Museum	1	1
Falun Dafa New York	1	1
Falun Municipality, Sweden	1	2
FalunArt.org	1	1
falunart.wizk.it	1	1
FalunCanada.net	2	7
FalunDafa.org	9	44
falungongtime.org	1	2
flowingdata.com	1	1
Free China Movement	1	1
Friends of Falun Dafa Radio	1	2
Friends of Falun Gong	1	5
hjclub.com	1	2
holyraymond.com	1	1
Human Rights Watch	2	4

<b>Organisation</b>	<b>Domains</b>	<b>URLs</b>
Independent Review	1	1
Mingjing.org	2	6
moikrug.ru (My Circle)	1	1
National Solar Observatory	1	1
NDTV.com	2	2
Omni talk forums	1	5
Open Magazine	1	1
osdir.com	2	5
phonecard.dajiyuan.com	1	3
Qianjinfang.net	1	1
Rick A. Ross Institute	1	2
ru-enlightenment.org	1	1
Sound of Hope Network	1	1
tradekey	1	1
Tripod	1	1
UCLA	1	1
ucnu.cn	1	1
Uncyclopedia	1	1
University of Amsterdam	1	1
University of Toronto	1	1
UpholdJustice.org	1	1
wenxuecity.com	1	1
Yu Ming	1	2
Zhengjian.org (PureInsight.org)	5	18
zhuanfalun.com	1	1
<b>Grand Total</b>	<b>83</b>	<b>186</b>

Table 14 Count of sometimes filtered Falun Gong URL site search results grouped by organisation

<b>Organisation</b>	<b>Domains</b>	<b>URLs</b>
64.71.164.40	1	1
AboutUs	1	1
Amazon	1	2
BBC	1	3
Britannica	1	2
CBC	1	3
CCTV	2	3
China.com	1	1
city.udn.com	1	1
Creaders.net	3	8
DisAbled Women's Network Ontario	1	2
Epoch Times	15	54
FGM TV	1	4
hardkingdom.com/freshrain	1	1
Minghui.org	31	131
MIT	1	2
Network54.com	1	1
ONI	2	2
ReligiousTolerance.org	1	1
renminbao.com	2	11
Sina.com	1	1
Stanford University	1	1
Time Magazine	1	1
Wikipedia	2	12
Yahoo Groups	1	1
<b>Grand Total</b>	<b>75</b>	<b>250</b>

Table 15 Count of always filtered Tibet URL site search results grouped by organisation

<b>Organisation</b> URLs	<b>Domains</b>	<b>URLs</b>
<b>Aboluowang</b>	<b>1</b>	<b>1</b>
<a href="http://www.aboluowang.com/">http://www.aboluowang.com/</a>	1	1
<b>BBC</b>	<b>1</b>	<b>7</b>
<a href="http://news.bbc.co.uk/">http://news.bbc.co.uk/</a>	1	7
<b>Boxun.com</b>	<b>1</b>	<b>1</b>
<a href="http://www.boxun.com/">http://www.boxun.com/</a>	1	1
<b>Central Tibetan Administration</b>	<b>1</b>	<b>4</b>
<a href="http://www.xizang-zhiye.org/">http://www.xizang-zhiye.org/</a>	1	4
<b>DWNews.com</b>	<b>1</b>	<b>1</b>
<a href="http://blog.dwnews.com/">http://blog.dwnews.com/</a>	1	1
<b>Human Rights Watch</b>	<b>1</b>	<b>2</b>
<a href="http://www.hrw.org/">http://www.hrw.org/</a>	1	2
<b>International Tibet Independence Movement</b>	<b>1</b>	<b>3</b>
<a href="http://www.rangzen.com/">http://www.rangzen.com/</a>	1	3
<b>Minzhuzhongguo.org (DemocraticChina.org)</b>	<b>1</b>	<b>1</b>
<a href="http://www.minzhuzhongguo.org/">http://www.minzhuzhongguo.org/</a>	1	1
<b>news.ju690.com</b>	<b>1</b>	<b>1</b>
<a href="http://news.ju690.com/">http://news.ju690.com/</a>	1	1
<b>renminbao.com</b>	<b>1</b>	<b>1</b>
<a href="http://renminbao.com/">http://renminbao.com/</a>	1	1
<b>SaveTibet.org</b>	<b>1</b>	<b>4</b>
<a href="http://www.savetibet.org/">http://www.savetibet.org/</a>	1	4
<b>Students for a Free Tibet</b>	<b>1</b>	<b>2</b>
<a href="http://www.studentsforafreetibet.org/">http://www.studentsforafreetibet.org/</a>	1	2
<b>tibetalk.com</b>	<b>1</b>	<b>2</b>
<a href="http://www.tibetalk.com/">http://www.tibetalk.com/</a>	1	2
<b>Grand Total</b>	<b>13</b>	<b>30</b>

Table 16 Count of sometimes filtered Tibet URL site search results grouped by organisation

<b>Organisation</b> URLs	<b>Domains</b>	<b>URLs</b>
<b>China.com</b>	<b>1</b>	<b>1</b>
http://www.china.com.cn/	1	1
<b>Epoch Times</b>	<b>2</b>	<b>6</b>
http://news.epochtimes.com/	1	4
http://www.epochtimes.com/	1	2
<b>MySpace</b>	<b>1</b>	<b>1</b>
http://vids.myspace.com/	1	1
<b>Wikipedia</b>	<b>1</b>	<b>4</b>
http://zh.wikipedia.org/	1	4
<b>Grand Total</b>	<b>5</b>	<b>12</b>

Table 17 Count of overblocked anti-Falun Gong/Tibet URLs grouped by organisation and domain

Organisation URLs	Unique URLs	All URLs
<b>202.84.17.11</b>	<b>1</b>	<b>6</b>
<a href="http://202.84.17.11/falun/">http://202.84.17.11/falun/</a>	1	6
<b>bbsland</b>	<b>1</b>	<b>5</b>
<a href="http://www3.bbsland.com/forums/general/messages/491649.html">http://www3.bbsland.com/forums/general/messages/491649.html</a>	1	5
<b>CCTV</b>	<b>3</b>	<b>4</b>
<a href="http://www.cctv.com.cn/specials/falungong/flgbaodao.html">http://www.cctv.com.cn/specials/falungong/flgbaodao.html</a>	1	1
<a href="http://www.cctv.com/news/special/zt1/XieJiaoFaLunGong/XieJiaoFaLunGong.html">http://www.cctv.com/news/special/zt1/XieJiaoFaLunGong/XieJiaoFaLunGong.html</a>	1	2
<a href="http://www.cctv.com/specials/falungong/flgbaodao.html">http://www.cctv.com/specials/falungong/flgbaodao.html</a>	1	1
<b>China.com</b>	<b>1</b>	<b>1</b>
<a href="http://www.china.com.cn/news/txt/2008-05/07/content_15091935.htm">http://www.china.com.cn/news/txt/2008-05/07/content_15091935.htm</a>	1	1
<b>Creaders.net</b>	<b>1</b>	<b>1</b>
<a href="http://bbs3.creaders.net/forums/general/messages/615014.html">http://bbs3.creaders.net/forums/general/messages/615014.html</a>	1	1
<b>Exposing the Falun Gong</b>	<b>1</b>	<b>5</b>
<a href="http://exposingthefalungong.org/">http://exposingthefalungong.org/</a>	1	5
<b>Mingjing.org</b>	<b>6</b>	<b>18</b>
<a href="http://www.mingjing.org.cn/dzlt/152.htm">http://www.mingjing.org.cn/dzlt/152.htm</a>	1	5
<a href="http://www.mingjing.org.cn/e-falun/cult/424.htm">http://www.mingjing.org.cn/e-falun/cult/424.htm</a>	1	3
<a href="http://www.mingjing.org.cn/zxxx/050914/02.htm">http://www.mingjing.org.cn/zxxx/050914/02.htm</a>	1	2
<a href="http://www.mingjing.org.cn/zxxx/050914/03.htm">http://www.mingjing.org.cn/zxxx/050914/03.htm</a>	1	1
<a href="http://www.mingjing.org.cn/zxxx/2020225/03.htm">http://www.mingjing.org.cn/zxxx/2020225/03.htm</a>	1	6
<a href="http://www.mingjing.org.cn/e-falun/cult/424.htm">www.mingjing.org.cn/e-falun/cult/424.htm</a>	1	1
<b>Network54.com</b>	<b>1</b>	<b>3</b>
<a href="http://www.network54.com/Forum/205697/thread/1162774893/last-1162774893/修炼“法楞功”造成心里障碍而导致死亡给我们的警醒">http://www.network54.com/Forum/205697/thread/1162774893/last-1162774893/修炼“法楞功”造成心里障碍而导致死亡给我们的警醒</a>	1	3
<b>Sina.com</b>	<b>1</b>	<b>1</b>
<a href="http://news.sina.com.cn/china/falungong/index.shtml">http://news.sina.com.cn/china/falungong/index.shtml</a>	1	1
<b>Grand Total</b>	<b>16</b>	<b>44</b>

Table 18 Count of overblocked unrelated URLs grouped by organisation

<b>Organisation</b> URLs	<b>Unique URLs</b>	<b>All URLs</b>
<b>BBC</b>	<b>2</b>	<b>7</b>
<a href="http://news.bbc.co.uk/chinese/simp/hi/newsid_4710000/newsid_4710300/4710359.stm">http://news.bbc.co.uk/chinese/simp/hi/newsid_4710000/newsid_4710300/4710359.stm</a>	1	1
<a href="http://news.bbc.co.uk/onthisday/hi/dates/stories/january/16/newsid_2530000/2530475.stm">http://news.bbc.co.uk/onthisday/hi/dates/stories/january/16/newsid_2530000/2530475.stm</a>	1	6
<b>Britannica</b>	<b>1</b>	<b>5</b>
<a href="http://www.britannica.com/eb/topic-276578/Hongzhi">http://www.britannica.com/eb/topic-276578/Hongzhi</a>	1	5
<b>China.com</b>	<b>1</b>	<b>1</b>
<a href="http://www.china.com.cn/news/txt/2008-01/09/content_9504599.htm">http://www.china.com.cn/news/txt/2008-01/09/content_9504599.htm</a>	1	1
<b>ChinaDaily.com</b>	<b>1</b>	<b>5</b>
<a href="http://www.chinadaily.com.cn/china/2008-01/30/content_6429758.htm">http://www.chinadaily.com.cn/china/2008-01/30/content_6429758.htm</a>	1	5
<b>Columbia University</b>	<b>1</b>	<b>1</b>
<a href="http://www.stat.columbia.edu/~cook/movabletype/archives/2008/04/rip_minghui_yu.html">http://www.stat.columbia.edu/~cook/movabletype/archives/2008/04/rip_minghui_yu.html</a>	1	1
<b>DAFA (Direct Aid for Africa)</b>	<b>1</b>	<b>7</b>
<a href="http://www.dafa.co.uk/">http://www.dafa.co.uk/</a>	1	7
<b>Falun Municipality, Sweden</b>	<b>2</b>	<b>10</b>
<a href="http://www.falun.se/">http://www.falun.se/</a>	1	3
<a href="http://www.falun.se/www/english.nsf">http://www.falun.se/www/english.nsf</a>	1	7
<b>flowingdata.com</b>	<b>1</b>	<b>6</b>
<a href="http://flowingdata.com/2008/04/12/reflecting-on-life-after-statistics-rip-minghui-yu/">http://flowingdata.com/2008/04/12/reflecting-on-life-after-statistics-rip-minghui-yu/</a>	1	6
<b>National Solar Observatory</b>	<b>1</b>	<b>3</b>
<a href="http://helios.tuc.noao.edu/">http://helios.tuc.noao.edu/</a>	1	3
<b>phonecard.dajiyuan.com</b>	<b>3</b>	<b>14</b>
<a href="http://phonecard.dajiyuan.com/product.asp?pid=46">http://phonecard.dajiyuan.com/product.asp?pid=46</a>	1	4
<a href="http://phonecard.dajiyuan.com/search.asp?callingfrom=USA">http://phonecard.dajiyuan.com/search.asp?callingfrom=USA</a>	1	3
<a href="http://phonecard.dajiyuan.com/tos.asp">http://phonecard.dajiyuan.com/tos.asp</a>	1	7
<b>Qianjinfang.net</b>	<b>1</b>	<b>2</b>
<a href="http://www.qianjinfang.net/cjsh/m/me/meifalun/meifalun.htm">http://www.qianjinfang.net/cjsh/m/me/meifalun/meifalun.htm</a>	1	2
<b>Tradekey</b>	<b>1</b>	<b>1</b>
<a href="http://www.tradekey.com/profile_view/uid/1309687/Hong-Zhi-Stone-Trade-Co-LTD.htm">http://www.tradekey.com/profile_view/uid/1309687/Hong-Zhi-Stone-Trade-Co-LTD.htm</a>	1	1
<b>Wikipedia</b>	<b>3</b>	<b>14</b>



<b>Organisation</b>	<b>Unique URLs</b>	<b>All URLs</b>
URLs		
<a href="http://en.wikipedia.org/wiki/Falun">http://en.wikipedia.org/wiki/Falun</a>	1	7
<a href="http://en.wikipedia.org/wiki/Hongzhi">http://en.wikipedia.org/wiki/Hongzhi</a>	1	2
<a href="http://en.wikipedia.org/wiki/Hongzhi_Emperor">http://en.wikipedia.org/wiki/Hongzhi_Emperor</a>	1	5
<b>Grand Total</b>	<b>19</b>	<b>76</b>

Table 19 Overblocked URLs that were simultaneously filtered and available in site search results

<b>Organisation</b>	<b>Unique URLs</b>
URLs	
<b>Britannica</b>	<b>1</b>
<a href="http://www.britannica.com/eb/topic-276578/Hongzhi">http://www.britannica.com/eb/topic-276578/Hongzhi</a>	1
<b>CCTV</b>	<b>3</b>
<a href="http://www.cctv.com.cn/specials/falungong/flgbaodao.html">http://www.cctv.com.cn/specials/falungong/flgbaodao.html</a>	1
<a href="http://www.cctv.com/news/special/zt1/XieJiaoFaLunGong/XieJiaoFaLunGong.html">http://www.cctv.com/news/special/zt1/XieJiaoFaLunGong/XieJiaoFaLunGong.html</a>	1
<a href="http://www.cctv.com/specials/falungong/flgbaodao.html">http://www.cctv.com/specials/falungong/flgbaodao.html</a>	1
<b>Columbia University</b>	<b>1</b>
<a href="http://www.stat.columbia.edu/~cook/movabletype/archives/2008/04/rip_minghui_yu.html">http://www.stat.columbia.edu/~cook/movabletype/archives/2008/04/rip_minghui_yu.html</a>	1
<b>DAFA (Direct Aid for Africa)</b>	<b>1</b>
<a href="http://www.dafa.co.uk/">http://www.dafa.co.uk/</a>	1
<b>flowingdata.com</b>	<b>1</b>
<a href="http://flowingdata.com/2008/04/12/reflecting-on-life-after-statistics-rip-minghui-yu/">http://flowingdata.com/2008/04/12/reflecting-on-life-after-statistics-rip-minghui-yu/</a>	1
<b>National Solar Observatory</b>	<b>1</b>
<a href="http://helios.tuc.noao.edu/">http://helios.tuc.noao.edu/</a>	1
<b>Qianjinfang.net</b>	<b>1</b>
<a href="http://www.qianjinfang.net/cjsh/m/me/meifalun/meifalun.htm">http://www.qianjinfang.net/cjsh/m/me/meifalun/meifalun.htm</a>	1
<b>Sina.com</b>	<b>1</b>
<a href="http://news.sina.com.cn/china/falungong/index.shtml">http://news.sina.com.cn/china/falungong/index.shtml</a>	1
<b>tradekey</b>	<b>1</b>
<a href="http://www.tradekey.com/profile_view/uid/1309687/Hong-Zhi-Stone-Trade-Co-LTD.htm">http://www.tradekey.com/profile_view/uid/1309687/Hong-Zhi-Stone-Trade-Co-LTD.htm</a>	1
<b>Wikipedia</b>	<b>2</b>
<a href="http://en.wikipedia.org/wiki/Hongzhi">http://en.wikipedia.org/wiki/Hongzhi</a>	1
<a href="http://en.wikipedia.org/wiki/Hongzhi_Emperor">http://en.wikipedia.org/wiki/Hongzhi_Emperor</a>	1
<b>Grand Total</b>	<b>13</b>

Table 20 Sensitive URLs that were simultaneously filtered and available in site search results

<b>Organisation URLs</b>	<b>Unique URL</b>
<b>Ascension Gateway</b>	<b>1</b>
<a href="http://www.ascensiongateway.com/quotes/li-hongzhi/index.htm">http://www.ascensiongateway.com/quotes/li-hongzhi/index.htm</a>	1
<b>Britannica</b>	<b>1</b>
<a href="http://www.britannica.com/eb/article-9342089/Li-Hongzhi">http://www.britannica.com/eb/article-9342089/Li-Hongzhi</a>	1
<b>holyraymond.com</b>	<b>1</b>
<a href="http://www.holyraymond.com/">http://www.holyraymond.com/</a>	1
<b>Independent Review</b>	<b>1</b>
<a href="http://duping.net/XHC/show.php?bbs=11">http://duping.net/XHC/show.php?bbs=11</a>	1
<b>osdir.com</b>	<b>1</b>
<a href="http://mlblog.osdir.com/network.freenet.general/2004-09/index.shtml">http://mlblog.osdir.com/network.freenet.general/2004-09/index.shtml</a>	1
<b>Wikipedia</b>	<b>1</b>
<a href="http://en.wikipedia.org/wiki/Li_Hongzhi">http://en.wikipedia.org/wiki/Li_Hongzhi</a>	1
<b>Grand Total</b>	<b>6</b>

## References

- Adedokun, O. A., & Burgess, W. D. (2011). Analysis of paired dichotomous data: A gentle introduction to the McNemar test in SPSS. *Journal of MultiDisciplinary Evaluation*, 8(17), 125-131. Retrieved from [http://journals.sfu.ca/jmde/index.php/jmde\\_1/article/download/336/337](http://journals.sfu.ca/jmde/index.php/jmde_1/article/download/336/337)
- Amnesty International. (2006). *Undermining freedom of expression in China: The role of Yahoo!, Microsoft and Google*. London: Amnesty International UK, the Human Rights Action Centre. Retrieved from <https://www.amnesty.org/en/documents/POL30/026/2006/en/>
- Ang, P. H. (2005). *Ordering chaos: Regulating the Internet*. Singapore: Thomson.
- Associated Press. (2006, June 7). Google 'compromised principles' over China. *The Guardian*. Retrieved from <http://www.guardian.co.uk/technology/2006/jun/07/news.searchengines>
- BBC. (2014, November 13). Tibet profile. BBC News. Retrieved from <http://www.bbc.com/news/world-asia-pacific-16689779>
- Battelle, J. (2005). *The search: How Google and its rivals rewrote the rules of business and transformed our culture*. New York: Portfolio.
- Berners-Lee, T. (1999). Glossary. *Weaving the Web*. Retrieved from <http://www.w3.org/People/Berners-Lee/Weaving/glossary.html>
- Burkeman, O., & Johnson, B. (2006, June 8). Google signals U-turn over Chinese site. *The Guardian*. Retrieved from <http://www.guardian.co.uk/technology/2006/jun/08/news.newmedia>
- Callaghan, T. (2014, December 22). Transparency Report: New numbers and a new look for government requests. Retrieved from <http://googlepublicpolicy.blogspot.ca/2014/12/transparency-report-new-numbers-and-new.html>

- Citizen Lab. (2007). *Everyone's guide to by-passing Internet censorship for citizens worldwide*. Citizen Lab, Munk School of Global Affairs, University of Toronto. Retrieved from [www.nartv.org/mirror/circ\\_guide.pdf](http://www.nartv.org/mirror/circ_guide.pdf)
- Cowley, S. (2006, January 27). Google CEO on censoring: "We did an evil scale". *MacWorld*. Retrieved from <http://www.macworld.com/article/1049166/googlechina.html>
- Crandall, J., Zinn, D., Byrd, M., Barr, E., & East, R. (2007). *ConceptDoppler: A weather tracker for Internet censorship*. Retrieved from [http://www.cs.unm.edu/~crandall/concept\\_doppler\\_ccs07.pdf](http://www.cs.unm.edu/~crandall/concept_doppler_ccs07.pdf)
- Damm, J., & Thomas, S. (2006). *Chinese cyberspaces: technological changes and political effects*. New York: Routledge.
- Drummond, D. (2010a, January 12). A new approach to China. Retrieved from <https://googleblog.blogspot.ca/2010/01/new-approach-to-china.html>
- Drummond, D. (2010b, March 22). A new approach to China: An update. Retrieved from <https://googleblog.blogspot.ca/2010/03/new-approach-to-china-update.html>
- Fallows, J. (2008). "The connection has been reset". *Atlantic Monthly*, 301, 64-69. Retrieved from <http://www.theatlantic.com/magazine/archive/2008/03/the-connection-has-been-reset/306650/>
- Faris, R., & Villeneuve, N. (2008). Measuring global Internet filtering. In R. Deibert (Ed.), *Access denied: The practice and policy of global Internet filtering* (pp. 5-27). Cambridge, Mass: The MIT Press. Retrieved from <http://access.opennet.net/wp-content/uploads/2011/12/accessdenied-chapter-1.pdf>
- Foster, W. A., & Goodman, S. E. (2000). *The diffusion of the Internet in China*. Stanford, CA: Stanford University: Center for International Security and Cooperation. Retrieved from <http://www.ciaonet.org/wps/fow01/FosterInternetChina2000.pdf>
- Google denies rumor of quitting. (2006, June 14). *SinoCast China IT Watch*. Retrieved from Factiva Database.

- Google. (n.d.). Google Company. Retrieved from <https://www.google.ca/about/company/>
- Google. (2007). Google Terms of Service. Retrieved from <https://www.google.com/policies/terms/archive/20070416/>
- Google. (2015). Search operators. Retrieved from <https://support.google.com/websearch/answer/2466433?hl=en>
- Gorman, G. E. (2005). China-bashing in the Internet censorship wars. *Online Information Review*, 29(5), 453-456. doi: [10.1108/14684520510628855](https://doi.org/10.1108/14684520510628855)
- Harwit, E., & Clark, D. (2001). Shaping the Internet in China: Evolution of political control over network infrastructure and content. *Asian Survey*, 41(3), 377-408.
- Herold, D. K., & De Seta, G. (2015). Through the looking glass: Twenty years of Chinese Internet research. *The Information Society*, 31(1), 68-82. doi: [10.1080/01972243.2014.976688](https://doi.org/10.1080/01972243.2014.976688)
- Human Rights Watch. (2006). *"Race to the bottom": corporate complicity in Chinese Internet censorship*. New York: Human Rights Watch. Retrieved from <http://www.hrw.org/reports/2006/china0806/china0806webwcover.pdf>
- Introvigne, M. (2015, September 21). Falun Gong: Religion. In *Encyclopaedia Britannica Online*. Retrieved from <http://www.britannica.com/topic/Falun-Gong>
- List of words censored by search engines in Mainland China. (2006, May 30). Retrieved from [http://en.wikipedia.org/wiki/List\\_of\\_words\\_blocked\\_by\\_search\\_engines\\_in\\_Mainland\\_China](http://en.wikipedia.org/wiki/List_of_words_blocked_by_search_engines_in_Mainland_China)
- Lumen. (2015, November 2). Retrieved from <https://cyber.law.harvard.edu/research/chillingeffects>
- Lumen. (n.d.-a). Canadian Jewish Congress demands removal of Zundelsite. Retrieved from <https://www.lumendatabase.org/notices/4678>

- Lumen. (n.d.-b). Complaint to Google France of racist/pro-Nazi material. Retrieved from <https://www.lumendatabase.org/notices/6870>
- Lumen. (n.d.-c). German complains to Google Groups of illegal Holocaust denial. Retrieved from <https://www.lumendatabase.org/notices/5991>
- McLaughlin, A. (2006, January 27). Google in China. Retrieved from <http://googleblog.blogspot.com/2006/01/google-in-china.html>
- OpenNet Initiative. (2006). *Internet filtering in China in 2004–2005: A country study*. Retrieved from <https://opennet.net/studies/china>
- OpenNet Initiative. (2008a, August 1). China Olympics update: more free for more people, but not free for all. Retrieved from <https://opennet.net/blog/2008/08/china-olympics-update-more-free-more-people-not-free-all>
- OpenNet Initiative. (2008b, December 19). China restores blocks on foreign websites. Retrieved from <https://opennet.net/blog/2008/12/china-restores-blocks-foreign-websites>
- OpenNet Initiative. (2008c). *Internet Filtering in China in 2006-2007*. Retrieved from <http://opennet.net/studies/china2007>
- OpenNet Initiative. (2008d, August 19). ONI analysis of Internet filtering during Beijing Olympic Games: Week 1. Retrieved from <https://opennet.net/blog/2008/08/oni-analysis-internet-filtering-during-beijing-olympic-games-week-1>
- OpenNet Initiative. (n.d.-a) About Filtering. Retrieved from <http://opennet.net/about-filtering>
- OpenNet Initiative. (n.d.-b) About ONI. Retrieved from <http://opennet.net/about-oni>
- Qiang, X. (2004, August 30). A list of censored words in Chinese cyberspace. *China Digital Times*. Retrieved from <http://chinadigitaltimes.net/2004/08/the-words-you-never-see-in-chinese-cyberspace/>

- Qiang, X. (2011). The battle for the Chinese Internet. *Journal of Democracy*, 22(2), 47-61.
- Qiu, J. (2010). A land without Google? *Nature*, 463(7284), 1012. doi: [10.1038/4631012a](https://doi.org/10.1038/4631012a)
- Roberts, J. B., & Roberts, E. A. (2009). *Freeing Tibet: 50 Years of Struggle, Resilience, and Hope*. New York: AMACOM. Retrieved from <http://www.ebscohost.com/ebook>
- Rothstein, J. (2006, June 9). Google co-founder says company is staying in China. *Reuters News*. Retrieved from Factiva database.
- Saich, T. (2004). *Governance and politics of China* (2nd ed.). New York: Palgrave Macmillan.
- Schrage, E. (2006, February 15). Testimony: The Internet in China. Retrieved from <http://googleblog.blogspot.com/2006/02/testimony-internet-in-china.html>
- Seeger, P., & Gabrielsson, A. (1968). Applicability of the Cochran Q test and the F test for statistical analysis of dichotomous data for dependant samples. *Psychological Bulletin*, 69(4), 269-277. doi: [10.1037/h0025667](https://doi.org/10.1037/h0025667)
- Smith, D. (2006, October 29). Google defiant over censorship in China. *The Observer*. Retrieved from <http://www.guardian.co.uk/technology/2006/oct/29/news.china>
- The Internet in China: A tool for freedom or suppression? Subcommittee on Africa, Global Human Rights, and International Operations and the Subcommittee on Asia and the Pacific of the Committee on International Relations, House of Representatives, United States Congress, 119th Cong. 2 (2006). Retrieved from [http://chrissmith.house.gov/uploadedfiles/gofa\\_feb\\_15\\_2006\\_hearing.pdf](http://chrissmith.house.gov/uploadedfiles/gofa_feb_15_2006_hearing.pdf)
- Tong, J. W. (2009). *Revenge of the Forbidden City: the suppression of the Falungong in China, 1999-2005*. New York: Oxford University Press.
- Tsui, L. (2001). Internet in China: Big mama is watching you (MA thesis). Retrieved from <http://www.lokman.nu/thesis/010717-thesis.pdf>

- Tsui, L. (2003). The panopticon as the antithesis of a space of freedom: Control and regulation of the Internet in China. *China Information*, 17(2), 65.
- Villeneuve, N. (2008). *Search Monitor Project: Toward a measure of transparency*. (Citizen Lab Occasional Paper No. 1). Retrieved from Citizen Lab Retrieved from <http://citizenlab.org/wp-content/uploads/2011/08/nartv-searchmonitor.pdf>
- Villeneuve, N. (2010, January 12). Google's new approach. Retrieved from <https://www.nartv.org/2010/01/12/googles-new-approach/>
- Vise, D. A., & Malseed, M. (2005). *The Google story*. New York: Bantam Dell.
- Wacker, G. (2003). The Internet and censorship in China. In C. R. Hughes, & G. Wacker (Eds.), *China and the Internet: Politics of the digital leap forward* (pp. 58-82). New York: RoutledgeCurzon.
- Washington Post. (2006, February 18). Keywords used to filter web content. *The Washington Post*. Retrieved from <http://www.washingtonpost.com/wp-dyn/content/article/2006/02/18/AR2006021800554.html>
- Wiener, S. (2011). Grass-mud horses to victory: The phonological constraints of subversive puns. Paper presented at the *Proceedings of the 23rd North American Conference on Chinese Linguistics*, 1, 156-172.
- Wojcicki, S. (2011, July). The eight pillars of innovation. Think with Google. Retrieved from <https://www.thinkwithgoogle.com/articles/8-pillars-of-innovation.html>
- Zittrain, J., & Elderman, B. (2003). *Empirical analysis of Internet filtering in China*. Retrieved from <http://cyber.law.harvard.edu/filtering/china/>