Control of Internet search engines in China
-- A study on Google and Baidu

By
Nan Wang

A thesis submitted in partial fulfilment of the requirement for the degree of Master of Computing at Unitec New Zealand, August 2008
Acknowledgements

This thesis would not have been possible without the patience and support of my supervisors Dr. Logan Muller, Dr. Xiaosong Li. I am also grateful to Dr. Donald Joyce.
Abstract

Internet is universal, open, information sharing media environment. The characteristics of Internet make it hard to be controlled. According to much literature, China has been believed to have setup its own Internet border to control the Internet. The Internet search engines provide global services to all users. When their services reach the Chinese Internet border, the Internet search engines have to apply Chinese Internet rules, regulations and laws. This research attempts to find out the real factors that influence control of Internet search engines in China, so it could help the readers and Internet search engines to build a picture of what and to what degree the Internet has been controlled in China. Methodologically, this research used real data that has been retrieved from real-life search engines. The sets of keywords which have been used were carefully selected based on a review of the literature. The data has been analyzed and findings have been compared with existing published literature. The final conclusions were not surprising. China is controlling its own Internet and it has been seen to be effective so far. The factors that influence control of Internet search engines were grouped and categorized into 3 different areas. They are political/religious factors which include news, blogs, BBS forums, opposition parties, political process in Taiwan, anti communist, independence movements, illegal religious movements; cultural/moral factors which include sex and homosexuality, health, violence; and economic factors. Among these factors, Chinese government has blocked most political/religious and cultural/moral factors related information and content.
Problem

There is legislation in China which impacts on the operation of the Internet in China, however, there is little published literature identifies the legislations impact on Internet search engines.

Research question

What are the factors that influence control of Internet search engines in China?

Keywords

Internet regulation, Internet law, Internet control, Internet filtering, Internet in China, search engine, keyword censorship, Google, Baidu
# Table of contents

Table of contents ................................................................................................................. 4  
List of tables ........................................................................................................................ 5  
Chapter 1: Introduction ....................................................................................................... 6  
  Purpose of research ....................................................................................................... 10  
Chapter 2: Literature review ............................................................................................. 11  
  Internet development in China ...................................................................................... 11  
  Internet control in China ............................................................................................... 15  
  Internet control in other countries ................................................................................. 25  
  Literature map ............................................................................................................... 29  
  Summary of literature review ....................................................................................... 30  
Chapter 3: Research methodology .................................................................................... 34  
  Sampling method .......................................................................................................... 38  
  Details of data gathering and compression method ...................................................... 39  
  Decision matrix ............................................................................................................. 43  
  Change management ..................................................................................................... 45  
Chapter 4: Discussions of primal search engines in China, Google and Baidu ............... 46  
Chapter 5: Findings and analysis ...................................................................................... 50  
  Discussions of each set of results .................................................................................. 53  
  Factors ........................................................................................................................... 58  
  Findings analysis ........................................................................................................... 61  
  Limitations of the research ............................................................................................ 63  
  Evading Internet control in China ................................................................................. 64  
Chapter 6: Conclusions ..................................................................................................... 65  
  Further research ............................................................................................................ 68  
References ......................................................................................................................... 69  
Appendix ........................................................................................................................... 74  
  Appendix 1 – Keywords list ......................................................................................... 74  
  Appendix 2 – Research results ...................................................................................... 82  
  Appendix 3 – Decision matrix ...................................................................................... 86
List of tables

Table 1 – Summary of research results (rank of keywords)……………………………………52
Table 2 – Keywords, ranks, category/area and factor………………………………………..58
Table 3 – Factors and sub categories/areas…………………………………………………..60
Chapter 1: Introduction

The government of the United States of America began to invest in the research and development of the Internet from the 1960s. Since then the Internet has grown rapidly and has become the most popular and widely used communication medium tool. “The Internet’s open, neutral architecture has proven to be an enormous engine for market innovation, economic growth, social discourse, and the free flow of ideas.” (Cerf, 2006)

Since the People’s Republic of China was founded in 1949, the content and distribution of the information has been monitored and controlled by the Chinese government. The appropriate content and distribution of information has been judged from the official Chinese Communist Party’s point of view. The Internet was brought to China in the mid 1990s, and Chinese Internet industry has been growing extremely fast since then. According to the latest survey from the China Internet Network Information Center (CNNIC), reported in “Internet development in China” on 30th June 2002, “the number of Internet users in China reached 45.8 million, a number which is more than the population of many nations in the world.” (Li, 2003) And by 2007, the number of Internet users in China has grown up to 137 million. (China Internet Network Information Centre, 2007) Comparing to other communication medium tools, the Internet could be considered as a more “democratic” tool in many places all over the world. Therefore, the introduction of the Internet in China has become a real challenge for the Chinese government.

Since China has reformed its market-oriented economy in 1979, China has become the world’s largest market, and many international business companies have moved into the Chinese trade market to gain more benefits. Although China has entered the World Trade Organization (WTO) in 2001, the Chinese Internet industry and telecommunication industry are still not free markets for foreign companies to enter. According to many past and current studies in the area of Internet control, the Chinese Internet control and Internet content filtering have been considered as the most complex and sophisticated around the world. (Lau, 2005; Li, 2003; Tsui, 2001) The Chinese government’s
techniques and methods which are used for Internet control issues include laws, regulations and many filtering techniques. When the world’s largest search engine company Google tried to get into the Chinese market, Google needed to address Internet control issues in the same way as many other Chinese domestic search engines and also needed to address some other factors such as intellectual property issues from both USA and China, and technical issues from both Chinese Internet industry and telecommunication industry.

“Google, Inc. (NASDAQ: GOOG and LSE: GGEA) is an American public corporation and search engine, first incorporated as a privately held company on 7 September 1998. The company had 9,378 full-time employees as of September 30, 2006 and is based in Mountain View, California.” (Wikipedia, 2006b) Google has grown extremely rapidly since it was founded in 1998, and the investment in the company has grown by millions and millions of dollars. Google’s core business service is based on its Internet search engines, and these search-oriented services include the web searching service, image searching service and Google News service. Google has also provided the Google Scholar searching service for the scholarly literature searching purpose. In the year 2004, the company launched Google earth and Gmail which is a free web mail service. And in early 2006, the company launched Google video which allows the users to search for available free videos. Since Google first started, it has already launched over 70 services and products. In October 2006, the company bought the most popular online video website “YouTube”, and the company plans to continue launching new products and services in the future.

With Google having been launched in many different regions and countries, Google has grown its business to include the whole world. When Google went to different markets of different regions and countries, there were many different factors to be considered by the company. These factors came from many different aspects such as culture, government control, laws, rules, regulations. The company also needed to be aware of its existing regional competitors. Some of the regional search engines are Accoona of China and US,
Baidu of China, Daum and Naver of Korea, In.gr of Greece, Rambler and Yandex of Russia, and Rediff of India.

So what happened in China? With over 1.3 billion people, China has been considered as the biggest market all over the world. As Google.com is a globally available search engine, so the Google users in China could also use the services through the Internet, but most of these users have struggled with Google.com services. Google.com has been down 10% of the time and most of the search services are not available. Therefore, to solve these problems, Google has decided to launch a Chinese local presence. Google.cn, which is the Google search engine for China, was launched in February 2006. The process of launching Google.cn has been considered as a course of action for Google. In order to launch Google.cn, Google had to agree to remove all the sensitive information within their search results. Also in consideration of the privacy of personal communication within China, Google.cn does not include Google’s most popular products and services such as Gmail and Blogger.

Baidu is another Internet search engine company. It was founded in 2000 and it is the most popular domestic search engine in China. Before Google moved into the Chinese market, Google invested in Baidu.com and owned 2.6% of its share until June 2006. At that time, Baidu.com was also called “Google of China”. “Baidu is a Chinese search engine which can search text and images. As of May 2006, it is fourth in Alexa's internet rankings. The Chinese word "Baidu" translates to "hundreds of times" in English. Baidu provides an index of over 740 million web pages, 80 million images and 10 million multimedia files.” (Wikipedia, 2006a). With the different copyright laws and intellectual property laws in China, Baidu.com’s most popular service is the MP3 searching service. It allows the users to search all music files with the formats such as MP3, WMA and SWF, these music files are mainly Chinese pop music files. Beside the text searching service and image service, Baidu.com also provides many different other products and services, these products and services include:
• “Baidu Zhidao -- A Q&A service
• Baidu Post -- A discussion group service based on keywords - upon entering a keyword, the user is taken to an existing discussion thread, or given the option of creating a new thread (similar to Google Base)
• Baidu Baike -- A wiki-like encyclopedia
• Baidu Space -- A free Social Networking platform
• Baidu Map -- Similar to Google Maps but detailed maps are available only for big cities”

(Wikipedia, 2006a)
Purpose of research

In order to access and stay in the Chinese Internet market and gain more competitive advantages, a study of control of Internet search engines in China has become important and necessary for both International and Chinese domestic Internet search engine companies. And to find the real impact of control of Internet search engines in China, this research takes the readers to China by searching various sets of keywords using Internet search engines. There is legislation in China which impacts on the operation of the Internet in China, however, there is little published literature identify the legislations impact on Internet search engines. Therefore, the following question has been asked to define the real factors that would affect Internet control and search engine companies’ interest:

What are the factors that influence control of Internet search engines in China?

In order to answer such a widely scoped main research question, a set of sub questions have been posed and these sub questions are:

- What are the current factors/issues for Internet development in China?
- What are the current factors/issues for Internet control in China?
- What are the current factors/issues for Internet control in other countries, such as USA and Vietnam?
- What are the differences in the results obtained using different Internet search engines with particular sets of non-sensitive and sensitive keywords?

These questions are important to ask because: to answer them will help to define the real Internet sensitive contents or censored keywords which have been controlled in China. Therefore, both international and Chinese domestic search engine companies should be aware of these issues in order to avoid unexpected problems and difficulties. These answers will also help researchers and scholars to understand the real situations of Internet control issues in China from both theoretical aspect and practical aspect.
Chapter 2: Literature review

For researchers based outside China, the literature that most retrieve has consistently been written from a Western perspective, the specific literature about the topic of Internet control in China and Internet control of search engines in China is very rare. This chapter will review the existing literature which discusses the Internet control in China and Internet control of search engines in China. The review of the literature will start from the current Internet development and usage in China, and then it will review what contents have been controlled by Chinese government and what strategies have been used to achieve its Internet control goal. Then the study of different and similar Internet control situations will be reviewed, the sample countries will be USA and Vietnam. Conducting this literature review will help to define the sensitive contents and censored keywords which have been controlled by Chinese government on the Internet search engines. It will also help to explore some hidden areas of Internet control in China. Therefore, a list of research keywords can be compiled in order to be used in this research to examine the true sensitive contents for Internet search engines in China.

Internet development in China

Brief history of the Internet in China

- China sent out first email (“Crossing the Great Wall to join the world.” “越过长城，走向世界”.) in September 1987.

- CAS (Chinese Academy of Sciences) carried out NCFC (National Computing and Networking Facility of China) project in October 1989.

- Prof. Qian Tianbai registered the Country Code Top-Level Domain .CN in November 1990.

- NCFC connected to the Internet on April 20, 1994. Since then, China has been officially recognized as a country with full functional Internet accessibility.

- In May 1994, the High Energy Physics Institute, CAS, set up China's first web server and made the first set of web pages
• In January 1995, Directorate General of Telecommunications began to provide Internet accessing services. From then on, the Internet began its commercialization process in China.

(Wang, 2003)

Since the first introduction of the Internet in China, Internet has become an important communication media for Chinese citizens. As stated by formal chairman of Internet society of China Prof. Hualin Qian, “Internet has achieved a rapid development that is head and shoulders above people’s imagination in terms of scales and functions. It has influenced human society from all aspects.” (Qian, 2004, P. 1) Internet has affected Chinese citizens in all fields, such as political, military, economic and culture.

The number of Internet users has been grown rapidly in China, in the year 2007, the number of Internet users in China has reached 137 million. The following figure from the China Internet Network Information Centre shows the growing number of Chinese Internet users during the past 7 years:

![Graph showing the growing number of Chinese Internet users](image)

(China Internet Network Information Centre, 2007, P. 24)

China has the world’s largest population, the number of Internet users will keep increasing in the future. “Total users: 137.0 million. The 95% confidence intervals for the estimate are 133.62 million to 140.38 million.” (China Internet Network Information Centre, 2007, P. 24) According to another report from Amnesty International, over 40% of these Internet users are based in Chinese most prosperous eastern cities like Beijing, Shanghai, Guangzhou and Shenzhen. (Amnesty International, 2002) With the economic
development in the western area of China in the recent years, there will be a large potential Chinese Internet market waiting to be discovered.

There are many different kinds of methods to access the Internet. According to the survey made by CNNIC (China Internet Network Information Centre), there were 27.1 million Chinese accessing Internet by leased lines (LAN users), 39 million Chinese accessing Internet by dial-up (including ISDN users), 90.7 million Chinese accessing Internet by broadband (including DSL and cable users), 17 million Chinese accessing Internet by mobile phones. (China Internet Network Information Centre, 2007, P. 5) And “besides computer and mobile phone, the number of the Internet users who simultaneously use other surfing devices (PDA or Information appliance) reached 1.78 million.” (China Internet Network Information Centre, 2007, P. 5) China has 59.4 million computer hosts, and has approximately 4,109,020 domain names. The following figure from the China Internet Network Information Centre shows the number and percentage of domain names in China by 2007:

<table>
<thead>
<tr>
<th>Table 2.3</th>
<th>Domain names by TLDs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
</tr>
<tr>
<td>CN</td>
<td>1,803,393</td>
</tr>
<tr>
<td>COM</td>
<td>1,939,343</td>
</tr>
<tr>
<td>NET</td>
<td>299,108</td>
</tr>
<tr>
<td>ORG</td>
<td>67,176</td>
</tr>
<tr>
<td>Total</td>
<td>4,109,020</td>
</tr>
</tbody>
</table>

(China Internet Network Information Centre, 2007, P. 6)

85% of Chinese Internet users prefer to obtain information from Internet and 47.4% of Chinese Internet users prefer Internet as their primary approach to obtain information. Search engine is the third most frequently used services and functions in China on Internet. There are 51.5% of Chinese Internet users would like to use search engines’ services and functions. (China Internet Network Information Centre, 2007, P. 13) Internet has become a part of Chinese citizen’s life and searching information through the search engines has also become an important part of surfing the Internet for Chinese Internet users. For both Chinese domestic and international search engine companies, the Chinese
market has the biggest potential increase and will allow them to earn huge profits from it. The four leading search engines in China are Yahoo, Baidu, Google and Sohu. And according to many surveys and studies, Baidu and Google have been rated to be the best search engines in China. The following figures from the report produced by CNNIC shows the popularity of these search engines being used in some of Chinese most prosperous eastern cities. “A search engine’s market share is determined by the number of users who use that particular search engine as their only or primary one, divided by the number of total users.” (China Internet Network Information Centre, 2005, P. 11)

![Search engine market share in Beijing](image)

(China Internet Network Information Centre, 2005, P. 12)

This study has been also reported by searchenginewatch.com, the search engine market share in Beijing in 2005 is Baidu: 51.5%, Google: 32.9%, Sohu: 4.6%, Sina: 4.0%, Yahoo: 3.7%, others: 3.3%. (Sherman, 2006) Along with other cities, it shows the similar results.

![Search engine market share in Shanghai](image)

(China Internet Network Information Centre, 2005, P. 13-14)
The Internet search engine is not only an online searching tool for the Internet users, it is also an online business environment for many business companies. According to the survey findings conducted by China IntelliConsulting Corporation (CIC), 86.3% of participant companies had put their advertisements on Baidu, 62.2% of participant companies had put their advertisements on Google, and 42.3% had put their advertisements on Yahoo. (China IntelliConsulting Corporation, 2007, P. 8) Therefore, the Internet search engine has provided a platform for businesses to present their services and products online and attract more customers.

**Internet control in China**

Openness and universality are the two very basic characteristics of Internet. Nowadays, Internet is universally open, information sharing media environment. This makes the information on Internet hard to be controlled by a particular individual, organization or country. “Truth is a powerful solvent. Stone walls melt before its relentless might. The Internet is one of the most powerful agents of freedom. It exposes truth to those who wish to see and hear it. It is no wonder that some governments and organizations fear the Internet and its ability to make the truth known.” (Cerf, 2003) The term of “Internet control” has different definitions in different countries. In western literature, it has been often defined as the functionalities of the Internet network. However in China, it does not only refer to how to control and filter the Internet network but also how to regulate any information and activities related to and on the Internet.

**Internet border of China**

After exploring the Internet development in China, we can find that people’s day to day life in China and economic development of China have been changed a lot by bringing in the Internet. And Chinese government is also implementing its own e-government project which is called “Government Online Project”. (Holliday & Yep, 2005) The goal of it will be setting up a paperless government in the future. At the same time, Chinese government also wants to tight on its control of the Internet. And now Chinese government’s Internet
filtering system has been considered as the most sophisticated and complex in the world. “The Chinese government tries to monitor individual use of the Internet through different means, including technical design, monitoring software, regulations, administrative and legal measures, and continuous political education.” (Sun, 2003) China is setting up its own Internet border, and it has been quite effective in that it allows the Chinese government to control and block most undesirable information. The government’s main filtering system takes place at the backbone level of the Internet network in China primarily, and the individual Internet service providers also implement their own filtering systems at their own level. While Chinese citizens are trying to view sensitive information, they will either get limited access to it or even could not reach it. The following diagram shows the Chinese government’s filtering system at the backbone level of the Internet network:

(Wu & Goldsmith, 2006)

Part of the Internet control in China is to filter and block the search results of particular sensitive topics or censored keywords on search engines. These search engines include both international search engines (Google.com) and as well as Chinese domestic search engines (Baidu.com). According to the Chinese government sponsored news agency, Xinhua news agency, the sensitive and censored targets means “superstitious, pornographic, violence-related, gambling and other harmful information”. (Lau, 2005,
P.3) However, the sensitive topics and censored keywords are also considered as the accessible materials which are undesirable to Chinese government. Such materials may include International news websites (including BBC News, VOA…), websites with dissident political objection, many Taiwanese commentary websites, and public encyclopedia (including Wikipedia…). When an Internet user is attempting to search for a sensitive keyword using these search engines in China, it will retrieve few or even no search results. For example, Google.cn will display the following message at the bottom of the search results page: “According to the local laws, regulations and policies, part of the searching result is not shown.” (“据当地法律法规和政策，部分搜索结果未予显示.”) The following diagram shows the process of retrieving information on Internet search engines:

And for further searching the keywords that are to be blocked, the user’s IP address or more potential third party machines can be blocked as well. “Once one connection has been blocked, the firewall makes further easy-to-evade attempts to block further connections from the same machine. This latter behavior can be leveraged into a denial-of-service attack on third-party machines.” (Clayton, Murdoch & Watson, 2006, P. 20-35)
Internal Internet control: laws and regulations

The Chinese government waited until 1995 to introduce the Internet, and there were no laws or regulations for the Internet during the first year. During the second year 1996, the Chinese government released its first main regulation which was a Temporary Regulation for the Management of Computer Information Network International Connection. “This regulation was formally announced on 1 February 1996 and verified on 20 May 1997”, (Tsui, 2001, P. 26) it is “the main regulation for governing the network infrastructure and international network connections”. (Tsui, 2001, P. 26) During this year 1996, China banned an estimated 100 websites by using its filtering system to prevent the distribution of sensitive information using the Internet. (Electronic Frontiers Australia, 2002) In an article “Great leap or long march: Some policy issues of the development of the Internet in China” (Wu, 1996, P. 699-711) which was published in 1996, it stated that the development of the Chinese Internet will be based on other countries’ Internet development experiences, and it will also be based on the Chinese’s own political, social and economic composition. (Wu, 1996, P. 699-711)

In 1997, Internet control law was added into Chinese Criminal Law. And the Chinese Criminal Law which was adopted in 1979 and amended in 1997 along with many other regulations allow the government to monitor and control the contents and distribution of the information throughout all kinds of media, this also affects the Internet. By the year 2002, there were over 60 Internet control laws, rules and regulations had been introduced in China, (Amnesty International, 2002) and there are far more existing today. These laws, rules and regulations cover the areas including state information security and information secrecy, and these laws, rules and regulations allow the government authorities to arrest and jail any individual or group of people for various offences that related to the Internet.

A more related catalog of Chinese key Internet control laws, rules, and regulations is contained in the article “Review of China's Internet regulations and domestic legislation.” (International Centre for Human Rights and Democratic Development, 2001) and from the government corresponding department “Chinese Internet Network Information Center.” (URL: http://www.farrall.org/china/home.html)
A report from Harvard Law School, covered Internet filtering in China and had come up with a scope of what had been blocked by Chinese Internet control. The access for some websites of dissident/democracy sites, education, news, health, government sites, Taiwanese and Tibetan sites, entertainment, religion had been denied. (Zittrain & Edelman, 2003) In this research, both literature review and findings will be analyzed and grouped into 3 categories and areas, and they will be political/religious, cultural/moral, and economic.

**Political/religious**

The control of the political and religious sensitive contents is the measurement for Communist Party of China to maintain their single party rule policy. The laws, rules and regulations help the Communist Party government to prevent and stop spreading any undesirable information. Such information may come from the unapproved political reformist, national separatist, counterrevolutionary or religious ideas. The laws and regulations do not only control where the information comes from, but also monitor and control the Chinese citizens to make sure they do not view such information. The websites and links affected by political and religious control normally include unapproved local media and international media, personal blogs and blog hosting service, foreign government and political parties. Chinese traditional religions including Buddhism, Taoism and Confucianism and revival religions including Chinese Islam and Christianity have been flourishing, Chinese citizens have been given greater religious freedom, but some religious issues are still existing and the information related to these issues is still banned and distribution controlled by Chinese government. And the foreign citizens who are willing to distribute the undesirable religious ideas will also be banned and controlled. The Falun Gong group has been defined as a social evil and all religious information related to this group has been banned.

The historical political and religious progress issues have been also monitored by Chinese Internet control system. People often study the events which happened in past and this will affect their behaviours in the future. “A historical perspective may further highlight
the contemporary relevance of the Internet. In Maoist China, political participation was
strictly guided by the state, and political dissent was a risky behaviour. In the reform
period beginning in 1978, mass political campaigns gradually receded from China’s
political scene while new, individualist modes of political participation appeared.” (Yang,
2003, P. 405-422) Yang also stated that it is very important for the civil society to
understand how and why the Internet has been shaped by politics. And Internet search
engine companies should be aware of any content related to these political implications
which may violate the Chinese Internet control system.

News agencies do not only broadcast local domestic news but also the international
events. Many foreign international news agencies which broadcast and critic the Chinese
political and religious processes have also been controlled and blocked by Chinese
government. Greatfirewallofchina.org website is used to test any websites which blocked
in China and it is “to be a watchdog and keep track of which and how many or how many
times sites are censored. Help to keep the censorship transparent. Each blocked website
will automatically be added to the great firewall on the homepage”.
(Greatfirewallofchina.org, 2007) According to the testing tool provided by
Greatfirewallofchina.org, the following websites and links are believed to be blocked for
political and religious sensitive reasons:

- BBC News (news.bbc.co.uk)
- China Times (www.chinatimes.com.tw)
- TVBS (www.tvbs.com.tw)
- Voice of America (www.voa.gov)
- WordPress Blogs (www.wordpress.com)
- Yahoo! Taiwan Blogs (tw.myblog.yahoo.com)
- Yahoo! Taiwan News (tw.news.yahoo.com)

**Cultural/moral**

Although there is much greater cultural freedom in China today, Chinese citizens are still
considered as having very strong traditional Chinese culture or Asian values. Along with
the rapid economic growth in China, more and more Western elements have been brought into Chinese citizens’ daily life, and some of them are considered as very positive to China and some are not. The sensitive information and contents have been justified as not suitable for Chinese culture and moral. Under the control of the Chinese government and many other organizations this undesirable information and content is prohibited from being accessed, such information includes any website and link that contains obscenity, pornography, criminal, and violence elements. Many online video stream services are also controlled and banned in China, because some of the films have been considered as containing pornography and violence scenarios. In order to play these films in China, the films have to delete the certain undesirable scenarios before being approved by the government department.

Health is another important aspect of cultural/moral issues. The contents related to medical process of public health in China have been monitored by Chinese Internet control system. “In May 2003 it was reported by the official Chinese News Agency, Xinhua, that over 100 people had been arrested for ‘spreading rumors’ or ‘false information’ through the Internet or mobile phone text messages about SARS. Little further information is available about these cases and it remains unclear exactly how many remain in detention.” (Amnesty International, 2004) This showed that not all the contents related to public health issues have been controlled but the false information and contents. Such information and contents may affect people with undesirable panic.

**Economic**

The Chinese Internet access providing industry is monopolized by government controlled Chinese local domestic companies. Currently, there are no other foreign companies which are allowed to provide such services. Chinese Internet access is “provided by nine state-licensed Internet access providers (IAPs), each of which has at least one connection to a foreign Internet backbone.” (OpenNet Initiative, 2005) The following figure shows the Chinese Internet access providers:
Nowadays, the control of Internet in China has been used not only for political, religious, culture and moral protection but also used for economic protection. The Internet information and contents of some foreign companies have been blocked to protect the Chinese local domestic business companies. Therefore, it could ensure a better opportunity for domestic economic growth. Any information and contents which would baffle the Chinese economy growth will be controlled and blocked. According to the article reported by Reporters Without Borders in March 2007, after publishing an article which contains the “risks of trading with China” contents, the website of the French organization Observatories International des Crises which produces the magazine on crises management methods for business was banned in China. (Reporters Without Borders, 2007)

**Internet police**

Who controls the Internet in China? According to the report from Reporters Without borders and Chinese Human Rights Defenders, the structure of Internet control in China is known as:

<table>
<thead>
<tr>
<th>Company Name</th>
<th>Audience</th>
<th>Network Name</th>
<th>Bandwidth (Mb/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>China Telecom</td>
<td>General public</td>
<td>CHINANET</td>
<td>462668</td>
</tr>
<tr>
<td>China Netcom/CNCNet</td>
<td>Broadband users</td>
<td>CHINA169</td>
<td>19087</td>
</tr>
<tr>
<td>China Science and Technology Network</td>
<td>Researchers</td>
<td>CSTNET</td>
<td>5275</td>
</tr>
<tr>
<td>China Unicom</td>
<td>Businesses</td>
<td>UNINET</td>
<td>1615</td>
</tr>
<tr>
<td>China Mobile</td>
<td></td>
<td>CMNET</td>
<td>1130</td>
</tr>
<tr>
<td>China Education and Research Network</td>
<td>Schools</td>
<td>CERNET</td>
<td>1022</td>
</tr>
<tr>
<td>China Int'l. Economic and Trade Net</td>
<td></td>
<td>CIETNET</td>
<td>2</td>
</tr>
<tr>
<td>China Great Wall Communications</td>
<td></td>
<td>CGWNET</td>
<td>Under construction</td>
</tr>
<tr>
<td>China Satcom</td>
<td></td>
<td>CSNET</td>
<td>Under construction</td>
</tr>
</tbody>
</table>

(OpenNet Initiative, 2005)
So have the real Internet police existed in China? The answer is “yes”. And the real Internet police have existed for a long time. The Internet police department has been set up in each police headquarters of cities. And they are monitoring the online servers, chatting rooms, and Internet cafes as well. They are the government authorities who will arrest and jail any individual or group of people for various offences that related to Internet. Beside real Internet police, there are other animated police being used in Chinese Internet control. Jingjing and Chacha (in Chinese, the term of characters “Jing Cha” means “police”) are the first two male and female cartoon Internet police officers from Shenzhen, China. They are first introduced to the public in January 2006, they are used to notify and educate the Chinese Internet users what is and is not legal to view and publish on Internet within China. Reported by China Digital Times and Director Chen from the Information Centre, Internet Security and Surveillance division, of Shenzhen Public Security Bureau said that “The Internet police have existed for a long time. This time we publish the image of Internet Police in the form of a cartoon, the purpose is to let all internet users know that the Internet is not a place beyond of law, the Internet Police will maintain order in all online behaviours.” (China Digital Times, 2006) When the
Chinese citizens view the Chinese web portals, the following animation character “Jingjing” and “Chacha” may be seen:

![The male cartoon Internet officer Jingjing.](image1)
![The female cartoon Internet officer Chacha.](image2)

(China Digital Times, 2006)

In year 2007, along with Shenzhen police, Beijing police also launched their version of animated cartoon Internet police officers. Now these offices are used by police on the Chinese web portals, they display the message about the Chinese Internet laws and regulations, and remind the Chinese Internet users to be legal citizens.
Internet control in other countries

“Different cultures have different attitudes towards the Internet and different national
governments have differing concepts of the role the Internet should play in society.” (Xie
& Jaeger, 2007, P. 1-15) In order to have a better understanding of the Chinese Internet
control system, it is necessary and important to explore the Internet control issues in other
countries. And this will provide a big picture which shows whether there are other
countries which have similar control or far different control of the Internet compared with
China. In this research, Vietnam and USA’s cases will be viewed and analyzed.

Vietnam

Vietnam (the Socialist Republic of Vietnam) is a Southeast Asian country. Vietnam was
unified in 1976. It is located at the easternmost corner of the Indochina Peninsula. The
population of Vietnam is over 85 million. It is bordered by China (north), Laos
(northwest), Cambodia (southwest) and the South China Sea (east coast line). The
Vietnam government is controlled by single party which is Vietnam Communist Party.
/Library of Congress - Federal Research Division, 2005/ According to government GDP
figures, Vietnam has been rated as the second fastest economic growth country in East
Asia and the fastest in Southeast Asia. Vietnam joined WTO in November 2006, its core
trading partners include USA, Japan, Australia, Southeast Asian developing countries and
Western European developed countries.

The Internet was first brought and introduced to Vietnam in 1996. During the first year,
Vietnam established its Internet links between USA, and Australia, with the speed of 64
KB per second for about 300 users. Since that time, the development in Vietnam has
spread rapidly and the number of Internet users has increased at 32.5% each year. By the
year 2004, the number of Internet users in Vietnam has reached 7.5 million which had
become the highest among the Southeast Asian countries, and the number of Internet
cafes in Vietnam largest city Ho Chi Minh City was over 1500. (Lam, 2006)
While the usage of Internet grows in Vietnam, various laws and regulations have been set up by Vietnam government to prevent the undesirable information. A department from Vietnam government which is called “Department General of Posts & Telecommunications (DGPT)” “coordinates and regulates management of the Internet for the authorities, while primary responsibility for its content is assigned to the Ministry of Culture and Information.” (Amnesty International, 2006) And Vietnam Internet “users are strictly prohibited from: Take advantage of the Internet to do hostile actions against the Socialist Republic of Vietnam or cause security unrest, violate morality and good customs and other laws and regulations.” (Amnesty International, 2006) Any critical opinion to the government and the distribution of any information which related to anti communism and anti government is sensitive. “Sensitive issues in Vietnam include corruption, ethnic unrest, and political opposition, including topics related to democracy or criticism of the state’s single-party system.” (OpenNet Initiative, 2006) The censored information and contents in Vietnam also focuses on political and culture/moral issues. In order to achieve government’s Internet control purpose, the Internet control laws and regulations have been used to repress Internet architecture and the filtering techniques and firewalls have been implemented at the ISP level. (OpenNet Initiative, 2006)

It seems that Vietnam would open its Internet to everyone. In the year 2001, the Vietnam prime minister announced that the Vietnam government would allow new ISPs to operate which will also include the privately owned ISPs as well. (Cerf, 2003) But according to the current studies, this has not taken place. The Internet in Vietnam is not very wildly spread and is under controlled by government. “The government has forbidden use of the Internet for political opposition, for actions against national sovereignty and security and violations of morality or the law.” (Cerf, 2003) “On 16 August 2002, Phan An Sa, deputy chief inspector at the culture and information ministry, called for access to subversive and pornographic material on the Internet to be blocked. He listed five kinds of Internet use he said was harmful to national security, including exchanges of anti-government material and use of the Internet to defraud people.” (Cerf, 2003)
USA

The USA (United States of America) is a federal constitutional republic comprising fifty states and a federal district. The country gained independence from Great Britain through the American Revolutionary War. Independence was declared on July 4, 1776 and finally was recognized on September 3, 1783. The country was considered a superpower after the World War II along with the Soviet Union, with the Soviet Union dissolved in 1991, the USA became the only superpower country in the post–Cold War era, it dominates international economic, political and cultural affairs, and its military forces are involved all over the world.

The concept of Internet control in USA is defined as the control of distributing or viewing information through the Internet. In USA, personal Internet access is not the matter to any technical control issues such as content control methods, but such methods are sometimes used in businesses, libraries, universities, and government department offices. Most of these actions are protected by laws and intellectual property regulations. USA uses the content control software to control the publication and distribution of the information on the Internet. These kinds of content control software are also known as censor ware or web filtering software. The use of content control software can be different and self customized according to different situations, such as target user groups, or specific locations. The decision of who should control the Internet has become an important issue not only for USA but for all the countries which connect to the Internet. Although we can say that the Internet users in USA have had much more freedom to surf the Internet and retrieve what they want from the Internet, the Internet users are still fighting against the Internet service providers and government to earn their Internet freedom. According to the report from Free Press in May 2006, (Scott, Cooper & Kenney, 2006) the Internet users have various reasons and facts to demand the Internet freedom.

People in USA have been given a great political freedom in their real life world. And the US government has also provided this great political freedom on the Internet as well. “The US government, for example, is one of a number of nations that has strongly embraced the political potential of the Internet, with the E-government Act driving as
much government information and as many services as possible into the online environment to promote political participation and interaction with the government.” (Xie & Jaeger, 2007, P. 1-15) There are many different ways and methods of different political parties and organizations to encourage people in USA to take part in political actions on the Internet.
Literature map

Control of Internet search engines in China

Internet development in China
- Wang, 2003
- Qian, 2004
- China Internet Network Information Centre, 2007
- Amnesty International, 2002
- China Internet Network Information Centre, 2005
- Sherman, 2006
- CIC, 2007

Internet border of China
- Cerf, 2003
- Holliday & Yep, 2005
- Sun, 2003
- Wu & Goldsmith, 2006
- Lau, 2005
- Clayton, Murdoch & Watson, 2006

Internal Internet control: laws and regulations
- Tsui, 2001
- Electronic Frontiers Australia, 2002
- Wu, 1996
- Amnesty International, 2002
- International Centre for Human Rights and Democratic Development, 2001
- Zittrain & Edelman, 2003
- Yang, 2003
- Greatfirewallofcchina.org, 2007
- Amnesty International, 2004
- OpenNet Initiative, 2005
- Reporters Without Borders, 2007
- China Digital Times, 2006

Internet control in Vietnam
- Xie & Jaeger, 2007
- Lam, 2006
- Amnesty International, 2006
- OpenNet Initiative, 2006
- Cerf, 2003

Internet control in other countries

Internet control in USA
- Scott, Cooper & Kenney, 2006
- Xie & Jaeger, 2007
Summary of literature review

The Internet content control is the main concern for Internet search engines when they try to launch their new services in different countries. And for many different reasons, in different countries, their Internet usage and control are different. In China, as world’s largest socialism country, the Chinese government has its own thoughts about Internet control for everyone in China. Given the many years of Internet development in China, China has developed its own “Internet border” to censor the contents which have been considered as not suitable for anyone to view. The current Internet development and usage situation can be retrieved from China Internet Network Information Centre, the statistical reports on the Internet development in China for each year contains various information about statistics of Internet users, statistics of computer hosts, statistics of WWW websites, statistics of international bandwidth and so on. But the censorship of Internet contents has not been mentioned at all. To research what contents have been controlled on the Internet in China, much literature related to western opinion has been reviewed because there is not much existing Chinese literature about this topic. There are several international groups monitoring the Internet control area of China, including “International Centre for Human Rights and Democratic Development” and “Reports without Borders and Chinese Human Rights Defenders”. These unofficial international groups have published reports that have reviewed the Chinese Internet control policies such as the laws, rules and regulations, and they also define the sensitive areas which have been controlled by Chinese government on the Internet.

After exploring different literature, there are many different websites and contents which have attracted the Chinese government’s attention on their Internet. They have been monitored and controlled to different levels and degrees. From the literature, the following main areas are defined as sensitive in China and this research will check if they truly are.

- Dissident/democracy websites
- Wiki/education websites(foreign universities websites such as Columbia, MIT)
- News websites
Control of Internet search engines in China

- Government websites
- Taiwanese and Tibetan websites
- Entertainment websites
- Religion websites
- Sexually explicit content websites
- Health content (such as famine china, AIDS china, disease china) websites
- Still many more contents have been controlled in China …

In this research, a list of keywords is selected according to the literature review. They are grouped into three defined categories and areas which have been believed to be controlled by Chinese government. They are political/religious, cultural/moral, and economic. To view the list of keywords, please refer to Appendix 1.

Internet is just a network which connects online computers together. It allows these computers to produce, share and distribute different information and contents within the Internet. After exploring different literatures, now it is time to summarise from this literature to draw some conclusions as to why the Chinese government wants to control its Internet. Referring back to the literature review, the websites and contents which have been controlled, mainly dissident/democracy sites, education, news, health and government sites, Taiwanese and Tibetan sites, entertainment and religion sites. (Zittrain & Edelman, 2003) Therefore, after analysing these different areas, both literature review and findings will be analysed and grouped into 3 categories and areas, and they are political/religious, cultural/moral, and economic.

- **Political/religious.** “An outside observer might not be able to offer a rational explanation for the puzzle, but the answer is quite straightforward: institutional legacy. Unlike western countries, the Chinese legal system has never paid more than lip service to the right of free speech. The demand from authorities usually precedes the need for privacy protection. In the Chinese political culture, there is no translation either to distinguish ‘censorship’ from ‘regulation’.” (Qiu, 2003) In western world, the political culture and participation normally means behaviours
which aim to shaping the current government’s policy, this involves the actions such as voting, protesting, serving in political parties and organizations. It is different situation in China, since China has the very narrow concept about its political culture and participation, and Chinese government does not want any harmful problems cause by this aspect. Therefore, any information, content and behaviour related to such political matters has been controlled.

- **Cultural/moral.** Cultural and moral values may have the same or different meanings to different countries and cultural groups. By exploring the literature about Internet control issues for different countries, it seems most countries and cultural groups are banning the undesirable moral contents such as violence materials. But as different cultural groups, Americans and Chinese are certainly showing their different culture values. This is just an example which has been examined in this research. Distributing sexual materials on any media has been considered as inappropriate in China. The Chinese government has set many regulations, rule and laws to prevent people from viewing inappropriate sexual materials. Internet as an emerging media in China is also affected by this situation. In USA, sexual material is allowed on the Internet. There are even online adult shops and adult websites gaining benefit from distributing sexual materials on the Internet.

- **Economic.** The Chinese Internet industry and market are still ruled by Chinese domestic companies. Therefore, China has the complete control at the ISP level. But as China joined WTO, the foreign companies are certainly hoping that they will have more freedom in the Chinese Internet industry and market. At the same time, any fake information and contents which may baffle the Chinese economy growth will be controlled and blocked.

What is missing in the existing literature and what will this research add to it? As mentioned before, most of literature about the topic Internet control in China is written from a western viewpoint, so what has happened in China? Is the real situation in China
the same as what has been written in this western viewpoint literature? This research will bring the readers to China and lead them to search different keywords in China, it will find out the real sensitive contents that have been controlled by Chinese government on the Internet search engines.

The technologies are changing rapidly with time. So the Internet and Internet search engines change also. The academic approved literature about the topic of Internet control in China has been limited, there are not many on different academic electronic databases and published books. Many information and opinions can be found on the news websites and public blogs and company conference records. This research could test the public ideas and find out which opinions are true and which fake ideas are.
Chapter 3: Research methodology

The research methodology is the set of methods used to examine the topic, gather and analyze information in order to answer the research question. This research will use a quantitative research methodology. The main research question and sub questions will be defined based on some concept studies (initial literature review). Then observation will be conducted by searching with different search engines (Google.com, Google.cn and Baidu.com) using different sets of keywords in order to provide statistically significant samples for each Internet search engine.

The purpose of designing such an observation is to have a better and easier control of the research situation, in order to evaluate the research findings based on the different search results. This research starts from the research question definition. The research question has been defined as how the Internet search engines are controlled in China. The main research question for this research is “What are the factors that influence control of Internet search engines in China?” and the sub questions are:

- What are the current factors/issues for Internet development in China?
- What are the current factors/issues for Internet control in China?
- What are the current factors/issues for Internet control in other countries, such as USA and Vietnam?
- What are the differences in the results obtained using different Internet search engines with particular sets of non-sensitive and sensitive keywords?

**Definition of sensitive:** As Hong Kong, Macau and Taiwan are the special administrative regions, they have their own legal systems, so the definition of sensitive within this research would not apply to these regions. The sensitive content is defined as any information that is limited and controlled within mainland China. This sensitive content can be categorized into different areas, and this research will only focus on these areas. The defined areas are political/religious, cultural/moral, and economic.
The initial literature review was carried out after research question definition. The research area explored included:

- Reviews of articles and papers on “the Internet development and control in China” which gathered some secondary information and answered the sub research questions “what are the current factors/issues for Internet development in China?” and “what are the current factors/issues for Internet control in China?”.  
- Reviews of articles and papers on “the Internet development and control in other countries” which gathered some secondary information on “what are the current factors/issues for Internet control in other countries?” During this research, one Asian developing country (Vietnam) and one developed country (USA) from the western world was studied. The reason of choosing these two countries is because Vietnam is an Asian socialistic country, so it might have Internet control issues similar to or different from China; USA and China are the two countries which have the largest number of Internet users all over the world, as a developed western country, USA might have Internet control issues which are quit different from China. By studying these two countries, it will finally answer sub research question “what are the current factors/issues for Internet control in other countries, such as USA and Vietnam?”.  

After the initial literature review, the topic and factors/issues on Internet search engines which has been controlled in China was able to be defined. Different sets of keywords were formed based on these different topics and also there was a list of random keywords formed to identify other possible factors/issues. The data required in this research was gathered by inserting different sets of keywords on different topics into the three Internet search engines which are Baidu.com, Google.com and Google.cn. The observed data was (a) the number of search results, and (b) the content of search results. Most sets of keywords will be formed by using online keyword creator. The key features of keyword creator is when the user enter a word or phrase, it will provide a list of words and phrases that are actually typed into the search engines by Internet users. By research through the Internet there are many available keyword creator tools, such as Wordtracker and Keyword Explorer from keyword-maker.com. Wordtracker is a privately owned
company which is based in London, UK. It is commercial software with the price $27.96/week, $55.93/month, and $279.64/year. (Wordtraker Keywords, 2007) It does provide the user with 1 week free trial. But as a low budget research, it was not suitable for this research. This research used Keyword Explorer from keyword-maker.com tool. Keyword Explorer allows the user to key in any type of keyword phase, and then it will generate a list of keywords that related to the keyword phase keyed in by the user. The Keyword Explorer also lists the frequency of each keyword searched by Internet users. For example with Keyword Explorer, you might type in word “pizza”, and then it will generate a list of words and phrases with that word in it that actually get used by Internet users. Such list may looks like:

- Pizza Hut
- Domino's Pizza
- Pizza dough
- Pizza recipes
- Pizza stone
- etc...

Five keywords from the list for each topic were chosen. Therefore, the non-sensitive keywords or sensitive keywords were categorized by different topics and these topics were related to different factors/issues which allowed drawing findings from.

This research also required different search techniques. The same non-sensitive and sensitive keywords were used from China and other different countries/regions (for this research it will be New Zealand). The data gathered at this stage was the quantitative data which was the number and the contents of the various different search results. Also the search engines were examined to determine the services they provided. There were some limitations doing this research, such as chances of search engine databases taking updates and changes during the research process, and this could affect the accuracy of the research. To avoid this situation, the data gathering stage of this research will be scheduled as short as possible by collecting it simultaneously from New Zealand and China. Therefore this required the researcher to gather the required data as quickly as possible.
How the keywords for this research were selected?

The keywords have been selected after the literature review and they have come from the reviewed literature studies.

- For example, in the comparison of keyword censorship chapter from “China journey to the heart of Internet censorship: Investigative report October 2007”, it defines some sensitive keywords which had been used in its research. These keywords were: “Communist Party of China’s conservative wing”, “Crusade against the Publicity Department”, “Wave of resignations by party members”, “Tiananmen massacre”, “Nine Commentaries on the Communist Party”. (Reporters Without Borders and Chinese Human Rights Defenders, 2007, P. 12-13)

- In “Internet filtering in China in 2004-2005: a country study”, it defines the Chinese Internet Content Regulation which includes “Regulation of Internet Users, Regulation of Content Providers, State Secrets Controls, News, Content Controls for Cybercafés, Extra-legal Content Controls.” (OpenNet Initiative, 2005) Lots of website links had been tested in their research and a list of which websites are accessible is summarized after their research.

Many different listed keywords may relate to the same censored area or certain event. After carefully selection and analysis, similar keywords which appeared in different literature have been put into same area. As the example shown above, the keywords of “anti communism”, “june 4”, “massacre china” are selected and will be used in this research process.

The reason of breaking the main keyword into sub keywords is: the main keyword is normally a single word or a short phrase and when it has been searched in search engines, there were huge number of results. The sub keywords are longer phrases and they normally contain richer information, so this will help to scope down the number of search results and make the results possible and easier to be analyzed.
Sampling method

As part of gathering the research data process, the sampling technique will be used. Sampling is “the process of using a small number of items or part of a larger population to make conclusions about the whole population.” (Zikmund, 2003) The first step of sampling is to define the target population. The observed data of this research will be the search results from the different Internet search engines. The population or the number of search results for a particular keyword could be in thousands or millions. The second step of sampling is to choose a sampling method. In this research, probability sampling technology and use random sampling method was chosen. Then the sampling frame and units were determined. The sampling frame is “the list of elements from which a sample may be drawn,” (Zikmund, 2003) and sampling unit is “a single element or group of elements subject to selection in the sample.” (Zikmund, 2003) The data required of these search results will include (a) the number of search results, (b) the content of search results. For this research, I will choose ten search results, these search results will be the 1st link and every fifth link. And the sampling units will include every working link within the sampling frame. Then the number of random working links within each keyword search will be calculated. The result was used to compare the number of working links within each of the different search engines, in order to define whether information related to such keyword was controlled or not.

All the gathered data was pre-edited and coded into an Excel spreadsheet. The data to be analyzed was the number and content of each search result found by searching through the different search engines using different sets of keywords. As there will be some screenshots to present the content of the search result, therefore a manual comparison method is required. Excel spreadsheet, manual comparison and statistical tests will be used in the data analysis process of this research.
Details of data gathering and compression method

Within each defined area (political/religious, cultural/moral, and economic), 20 keywords were chosen. Each keyword was entered into the Keyword Explorer and the first top 10 sub keywords from the list were picked, these sub keywords were used to insert into the different search engines to get different search results. To insert keyword into search engines, the same keyword was inserted into Google.com and Google.cn at almost the same time in different locations, for this research it will be New Zealand and China. The reason this was done was:

- When the user searches in New Zealand on Google.com, it should be considered to have no control or block issues at all.
- When the user searches in New Zealand on Google.cn, it should be considered to have been controlled or blocked by both China gateway and Google.cn local database.
- When the user searches in China on Google.com, it should be considered to have been only controlled or blocked by China gateway.
- When the user searches in China on Google.cn, it should be considered to have only been controlled or blocked by Google.cn local database.
Control of Internet search engines in China

Process of data gathering flow chart

New Zealand

Start

Open search engines:
www.google.com
www.google.cn

Me (Ray)

Pick a keyword

Search picked keyword on search engines

Send over keyword

Retrieve Total NO. of search results

Key in Total NO. of search results to Excel spreadsheet

Sampling search results *

Retrieve Sampling result

Key in Sampling result to Excel spreadsheet

Pick another keyword and start again

China

Start

Communicate by using MSN

My brother

Open search engines:
www.google.com
www.google.cn

Search picked keyword on search engines

Retrieve Total NO. of search results

Sampling search results *

Retrieve another keyword and start again

* Expand to detailed flow chart
Expand detailed flow chart for sampling search results process

1. Sampling search results
2. Retrieved all search results
3. Examine the 1\textsuperscript{st} link and every 5\textsuperscript{th} link
4. Is it a bad link?
   - Yes: Repeat and examine 20 sampling links
   - No: Record bad link

Sampling result = \((20 – \text{number of bad links}) / 20 \times 100\%\)

Number of working links = \(\text{total number of links} \times \left(\frac{20 – \text{number of bad links}}{20}\right)\%\)

To calculate the number to working links, 20 links among the search results were taken and they were the first link and every 5\textsuperscript{th} link. So the number of working links = total links \(\times \left(\frac{20 – \text{number of bad links}}{20}\right)\%\). Then code and compare the number of working links by using Excel. The basic principle of compression is shown as below:
Make different statistics comparison of the following statistics:

- The total number of search results and number of working links retrieved from Google.com in New Zealand.
- The total number of search results and number of working links retrieved from Google.com in China.
- The total number of search results and number of working links retrieved from Google.cn in New Zealand.
- The total number of search results and number of working links retrieved from Google.cn in China.

In order to make the comparison easier to be analyzed, these 4 sets of results will be put together into 2 groups as results retrieved in New Zealand and China.

The result from this comparison provided a top down list of the most sensitive keyword to least sensitive keyword. At the same time the services provided by Google.com, Google.cn and Baidu.com were also compared. Analyzing these findings enabled the main research question to be answered. The analysis process will be explained next.
Decision matrix

The decision matrix is a quantitative method of describing and ranking the gathered information in research. The decision matrix assigns different weight to different options, entities and criteria, and calculates the scores in order to ranking the options and entities. The complete decision matrix document has been attached as Appendix 3. The following picture shows the screenshot example of decision matrix and further explanation will be based on this example.

(Screenshot example of decision matrix, refer to Appendix 3)
The process of decision matrix

The criteria can be also called decision model in a decision matrix. In this research, I developed 2 groups of criteria:

- Whether the total number of search results and number of working links retrieved in New Zealand $\geq <$ (greater than or equal to or less than) the total number of search results and number of working links retrieved in China
- Sampling result is between different ranges
  - $95\% - 100\%$
  - $85\% - 95\%$
  - $65\% - 85\%$
  - less than $65\%$

In order to make the analysis process easier and attractive, the sampling results will be color coded. All keywords are given different colors.

- Unmarked keyword – Have not been affected at all.
- Green keyword – Sampling result is above $85\%$, and search results retrieved from China are less than search results retrieved from New Zealand.
- Yellow keyword – Sampling result is at the range of $65\%$ to $85\%$, and search results retrieved from China are less than search results retrieved from New Zealand.
- Red keyword – Sampling result is blow $65\%$, and search results retrieved from China are less than search results retrieved from New Zealand.

Then the weight has been assigned to each criteria based on its importance in the final research decision making process.

- The weight has been assigned to whether the total number of search results and number of working links retrieved in New Zealand $\geq <$ the total number of search results and number of working links retrieved in China has been assigned as yes or no. it will allow me to compare the difference between the same keywords in different locations.
- The weight has been assigned to different sampling result range is from 1 to 4. 1 means the internet control has less impact on the keyword and 4 means the
keyword has been impacted by the Internet control. This allowed me rate the score for different keywords. Then for each sub keyword a score has been rated according to its sampling result, and the score for each set of keyword is the sum of its sub keywords’ score. Then the rank will be calculated by score of keyword divide by number of sub keywords. For example, the total score of keyword “democracy china” is 4+1+2+4+1+4+4+2+4=26, the count of number of sub keywords is 9, therefore the rank of keyword “democracy china” is 26/9=2.8889.

The final findings from this research were formed after processing and analyzing the gathered data. This allowed answer to be provided for the sub research question “what are the differences in the results obtained using different Internet search engines with particular sets of non-sensitive and sensitive keywords?”. The conclusions and recommendations focus on and review the sub research questions and finally answer the main research question.

**Change management**

During the research period, change occurred. As stated in the previous chapters, I was going to use the Keyword Explorer from keyword-maker.com to generate my keywords, but when I started my keywords generation work, this tool had become commercial software which was not free to use any more. So with the comparison between the other tools, I decided to use the keywords generation function on the Enteryourkeyword.com. ([http://www.enteryourkeyword.com/index.php](http://www.enteryourkeyword.com/index.php)) The website provides similar functions to the Keyword Explorer from keyword-maker.com, it allows the user to key in any keywords and will generate a list of keywords related to it.
Chapter 4: Discussions of primal search engines in China, Google and Baidu

The Internet is considered as the open environment by most countries around the world. But in China, it has different meanings. The Chinese government has repeatedly announced that the Internet in China should be legally deployed. And by now, China has the world's most complicated regulations and laws for Internet. These regulations and laws are not only for Chinese Internet users but for both Chinese local and foreign Internet investors as well. Therefore, Internet search engines are also affected by these control activities. They are facing a heavy set of Internet control regulations and laws. Many of the laws are directly controlling the content distribution on the Internet and are affecting the search results of different Internet search engines directly as well. Such laws like: “Decision of the Standing Committee of the National People's Congress on Maintaining Internet Security (2000); Measures for Managing Internet Information Services (2000); Provisional Rules for the Administration of the Operation of News Publication Services by Web Sites (2000); Rules for the Administration of Internet Bulletin Board System Services (2000); Rules for the Administration of Computer & Internet Bulletin Board System Services in the Colleges (2001); Interim Provisions on the Administration of Internet Publication (2002).” (Li, 2003) “The English versions of these statutes could be available at www.isinolaw.com or www.lawinfochina.com.” (Li, 2003) So it is very important for Internet search engine companies to study and understand these issues, especially the foreign Internet search engine companies. Once they decide to run their business in China, at the very beginning they should do such kind work to avoid Chinese Internet control barrier.

As the research was taking place, the figures were changing rapidly. In the latest version of the CNNIC report, there are more and more Chinese citizens aged 6 and above surfing on the Internet now, and it seems that the growth of Internet users’ number will overtake the US shortly. Therefore, China will soon become the world's largest Internet market, it has remarkable meaning for the Internet search engines. How to attract Chinese Internet users to use their services has become more and more important.
Providing more and more useful services and tools beside the fully functioned online searching abilities will attract more users. As the Chinese leading local search engine, Baidu has launched more and more new services in these last few years to increase its competitive advantages in order to gain more users. For instance, a new voice based search service called "yellow page" has been launched, it provides a call in function and a real operator will help you to do the search. A journalist from Search Engine Journal reports that “You can call (010) 1161 6690 and ask the operator to search for information. Your correspondent called and asked for a Middle Eastern restaurant in Beijing. The operator took about one minute to provide the phone number and address of 1001 Nights, which is in fact the capital’s most popular Arab restaurant. I then asked the operator to find Hu Jintao’s (note: the Chinese President) email address. She was amused but had no luck finding it.” (Search Engine Journal, 2007)

Google services and tools

<table>
<thead>
<tr>
<th>Search</th>
<th>Explore and innovate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emails</td>
<td>Code</td>
</tr>
<tr>
<td>Image search for images on the web</td>
<td>Download APIs and open source code</td>
</tr>
<tr>
<td>Blog Search</td>
<td>Custom Search News</td>
</tr>
<tr>
<td>Find blogs on your favorite topics</td>
<td>Create a customized search experience for your community</td>
</tr>
<tr>
<td>Book Search</td>
<td>Labs</td>
</tr>
<tr>
<td>Search the full text of books</td>
<td>Explore Google's technology playground</td>
</tr>
<tr>
<td>Checkout</td>
<td>Communicate, show &amp; share</td>
</tr>
<tr>
<td>Complete online purchases more quickly and securely</td>
<td>Blogger</td>
</tr>
<tr>
<td>Desktop</td>
<td>Share your life online with a blog – it’s fast, easy and free</td>
</tr>
<tr>
<td>Search and personalize your computer</td>
<td>Calendar</td>
</tr>
<tr>
<td>Earth</td>
<td>Organize your schedule and share events with friends</td>
</tr>
<tr>
<td>Explore the world from your computer</td>
<td>Docs</td>
</tr>
<tr>
<td>Finance</td>
<td>Create and share your projects online and access them from anywhere</td>
</tr>
<tr>
<td>Business info, news, and interactive charts</td>
<td>Gmail</td>
</tr>
<tr>
<td>Google News</td>
<td>Fast, searchable email with less spam</td>
</tr>
<tr>
<td>Find and connect with businesses from your phone, for free</td>
<td>Orkut</td>
</tr>
<tr>
<td>Google</td>
<td>Create mailing lists and discussion groups</td>
</tr>
<tr>
<td>Add news, games and more to the Google homepage</td>
<td>Picasa</td>
</tr>
<tr>
<td>Special Search</td>
<td>Find new people and stay in touch with friends</td>
</tr>
<tr>
<td>Search within specific topics</td>
<td>Reader</td>
</tr>
<tr>
<td>Scholar</td>
<td>Get all your blogs and news feeds fast</td>
</tr>
<tr>
<td>Add a search box to your browser</td>
<td>SketchUp</td>
</tr>
<tr>
<td></td>
<td>Build 3D models quickly and easily</td>
</tr>
<tr>
<td></td>
<td>Talk</td>
</tr>
<tr>
<td></td>
<td>Fill and call your friends through your computer</td>
</tr>
<tr>
<td></td>
<td>Translate</td>
</tr>
<tr>
<td></td>
<td>View web pages in other languages</td>
</tr>
<tr>
<td></td>
<td>YouTube</td>
</tr>
<tr>
<td></td>
<td>Watch, upload and share videos</td>
</tr>
</tbody>
</table>

(Google, 2007)
Baidu services and tools

But similar to the content of search results, the use of services and tools are also under control by the Chinese government. Some of the services and tools on Google are available outside China, but maybe not available inside China. And by comparing the available services and tools provided on Google.com and Google.cn, there are less services and tools available on Google.cn than Google.com. These facts are driven by the Chinese Internet control issues. For example, there is no blog and Gmail service on Google.cn. When Google first start its business in China, it has tried to launch these popular services and tools, but they were removed because of the Chinese Internet privacy and content control matters.

(Baidu, 2007)
Although the search engines are having many problems when they facing the Chinese Internet control issues, there are still benefits that could be gained from such control issues. As the differences between the copyright laws and intellectual property laws in different countries, in China, the search engines are allowed to share Chinese mp3 music. This allows Baidu to provide its most popular service and tool for MP3 music searches, and this directly attract more and more younger Internet users to it.
Chapter 5: Findings and analysis

The actual research results and decision matrix analysis results of this research have been recorded in the Excel spreadsheet, and have been attached with this thesis as Appendix 2 and 3. The findings and conclusions of this research are based on these research results and decision matrix analysis results.

In this research, the literature review was used to identify key areas of internet control in China. From this literature review, typical keywords were identified to implement my research method. The keywords were then added to a keyword generator to expand this list. The keyword list and generated sub keywords has been attached as Appendix 1. The different sets of keywords were color coded as categories and themes emerged from the analysis process as described in the methodology chapter. The color coding used is:

- Unmarked keyword – Have not been affected at all.
- Green keyword – Sampling result is above 85%, and search results retrieved from China are less than search results retrieved from New Zealand.
- Yellow keyword – Sampling result is at the range of 65% to 85%, and search results retrieved from China are less than search results retrieved from New Zealand.
- Red keyword – Sampling result is blow 65%, and search results retrieved from China are less than search results retrieved from New Zealand.

How to rank the keywords?

In order to sort the search result, a scoring and ranking method was applied. This is described below:

- Unmarked keyword – score 1
- Green keyword – score 2
- Yellow keyword – score 3
- Red keyword – score 4

Add all the score for each set of keyword and then divide by the number of sub keywords under each set of keyword. For example, refer to the decision matrix excel spreadsheet.
and decision matrix chapter of this research, the 1st set of keyword “democracy china”. There are 9 sub keywords under it, and total score of them added up is 26. So the rank of this keyword is total score divide by number of sub keywords (26 / 9 = 2.8889). The rank will be at the range of 4.0000 to 1.0000, the higher rank (closer to 4.0000) means the keyword has been affected by the Chinese Internet control issues and the lower rank (closer to 1.0000) means the keyword may have been less affected or not affected by the Chinese Internet control issues.

**The reason of ranking the keywords:**

The rank calculates the rank value in a group of statistics. It allows giving a group of statistics some kind of ordering, therefore, the statistics could be analyzed according to such ordering. The use of top down list is a strategy in information processing and analysis. With the group of statistics has been ranked, a top down list will provide an accurate and easy way in information processing and analysis. In this research, a rank range of 1.000 to 4.000 has been decided, this comes from the design of using the decision matrix as explained previously.

**Top down list of keywords**

The rank of each set of keyword has been calculated after following the data gathering process and decision matrix process introduced in research methodology chapter. The research results have been ordered into a top down list. The following table shows the sensitivity of each keyword used.
<table>
<thead>
<tr>
<th>Keyword</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>falun gong</td>
<td>4.0000</td>
</tr>
<tr>
<td>Sequence search of “falun”, “lungong”, “lunda”, “dafa”</td>
<td>4.0000</td>
</tr>
<tr>
<td>massacre china</td>
<td>4.0000</td>
</tr>
<tr>
<td>minghui</td>
<td>4.0000</td>
</tr>
<tr>
<td>sex china</td>
<td>3.5455</td>
</tr>
<tr>
<td>playboy</td>
<td>3.5000</td>
</tr>
<tr>
<td>anti communist</td>
<td>3.4000</td>
</tr>
<tr>
<td>aids china</td>
<td>3.3333</td>
</tr>
<tr>
<td>protect rights</td>
<td>3.3000</td>
</tr>
<tr>
<td>dictatorship</td>
<td>3.1818</td>
</tr>
<tr>
<td>june 4</td>
<td>3.0833</td>
</tr>
<tr>
<td>genocide china</td>
<td>3.0000</td>
</tr>
<tr>
<td>hongzhi</td>
<td>3.0000</td>
</tr>
<tr>
<td>democracy china</td>
<td>2.8889</td>
</tr>
<tr>
<td>voice of America</td>
<td>2.8000</td>
</tr>
<tr>
<td>tvbs</td>
<td>2.6000</td>
</tr>
<tr>
<td>dalai lama</td>
<td>2.3636</td>
</tr>
<tr>
<td>independence china</td>
<td>2.3333</td>
</tr>
<tr>
<td>treason</td>
<td>2.2000</td>
</tr>
<tr>
<td>BBC news</td>
<td>2.0909</td>
</tr>
<tr>
<td>human rights</td>
<td>1.9000</td>
</tr>
<tr>
<td>humanity</td>
<td>1.8000</td>
</tr>
<tr>
<td>coup</td>
<td>1.0000</td>
</tr>
<tr>
<td>disciple</td>
<td>1.0000</td>
</tr>
<tr>
<td>oppression</td>
<td>1.0000</td>
</tr>
<tr>
<td>overthrow</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Table 1 – Summary of research results (rank of keywords)
Discussions of each set of results

Following is a discussion that expands on each of the ranking obtained from each of the main keywords researched. The ranking is the result of the process of scoring described at the beginning of this chapter.

Keyword list

- falun gong (rank: 4.0000)

IP address blocked when searching this term of keyword. While searching the keyword “falun gong”, it did not only block the search results but also the IP address that was used to search. The fact is that it first showed the browser can not display the page and then it blocked all other searches as well. For example, if searching “falun gong” on Google.com or Google.cn in China, it shows: (if the user uses the IE7.0, the result is shown as below)

And then if going back and trying to search any keyword again, it blocks all the results and shows the above same result again. To continue the search, the action of restarting the router or modem has to be undertaken in order to refresh IP and start fresh again.
• Sequence search of “falun”, “lungong”, “lunda”, “dafa” (rank: 4.0000)
The Internet sequence block technique has been used to block such keywords related to “falun gong”. The same situation occurs as searching “falun gong”.

• massacre china (rank: 4.0000)
There is only one keyword retrieved from Enteryourkeyword.com. This retrieved keyword has a great impact of the topic “massacre china”. And while searching it, less than 10 results have been retrieved both in New Zealand and China. And more than half of them have been blocked while search in China, around 60% of results were accessible while searching in New Zealand.

• minghui (rank: 4.0000)
The keyword “minghui” is the Chinese name of the falun gong website. The same situation occurs as searching “falun gong”.

• sex china (rank: 3.5455)
Most of the search results retrieved while searching this set of keywords were rated adult only contents. The results included mostly the porn graphics, videos, websites, webcams and online stores etc. Most of these results are working while searching inside New Zealand. According to the law of China, the behaviors of producing and distributing such contents through any media are illegal. So it is no surprise that most of these results have been blocked, but there are still over 40% of such results working while searching inside China.

• playboy (rank: 3.5000)
The same situation occurs as searching “sex china”. And at this time, there are still over 45% of such results working while searching inside China

• anti communist (rank: 3.4000)
The set of keywords retrieved from Enteryourkeyword.com for “anti communist” are not related to China specifically. There is no result retrieved from searching
the first keyword and there is only less than 5 results retrieved for the second keyword, and more than 70% of the results had been blocked. Therefore, “anti communist” has not only censored by Chinese government but also has very limited retrievable contents in New Zealand as well.

- aids china (rank: 3.3333)
  There are only three keywords retrieved from Enteryourkeyword.com. There is no result retrieved from searching “aids china cultural eating in politics rice southwest spring” and there are 5% to 45% of the rest search results have been blocked.

- protect rights (rank: 3.3000)
  While searching the keywords, 3 out of 10 keywords retrieved 0 search result. And there are 20% to 75% of the rest search results have been blocked.

- dictatorship (rank: 3.1818)
  Most of keywords retrieved from Enteryourkeyword.com are related to Cuba and Argentina. Most of the search results are accessible, but there are 15% to 45% of the search results that have been blocked.

- june 4 (rank: 3.0833)
  The term of keyword “june 4” is related to Tiananmen massacre. While searching the keywords related to Chinese Tiananmen massacre in 1989, over 40% of the search results have been blocked.

- genocide china (rank: 3.0000)
  There is only one keyword retrieved from Enteryourkeyword.com. And most of the search results are accessible, but there are many bad links in these search results, either the servers of the websites are down or certain content had been removed from the servers.
• hongzhi (rank: 3.0000)
The term of keyword “hongzhi” is the name of Falun Gong leader. There is a surprise that it does not have the same situation as search “Falun Gong”. But the most of search results retrieved are the news and government announcements, in these search results “hongzhi” has been defined as criminal.

• democracy china (rank: 2.8889)
While searching the keywords, 4 out of 9 keywords retrieved 0 search result. And among the rest search results, there are 5% to 50% of them have been blocked.

• voice of America (rank: 2.8000)
The most of the search results are accessible, but while searching some keywords, there are 60% of the search results that have been blocked. Therefore, the Chinese Internet control has an impact on foreign media.

• tvbs (rank: 2.6000)
The term of keyword “tvbs” is the name of Taiwan TV channel and network. 40% to 60% of the retrieved search results from searching the keywords “tvbs Taiwan” and “tvbs news Taiwan” have been blocked. Therefore, the Chinese Internet control has impact on contents related to Taiwan news and TV critics.

• dalai lama (rank: 2.3636)
Most of the search results are accessible. 5% to 45% of the search results have been blocked.

• independence china (rank: 2.3333)
There are only three keywords retrieved from Enteryourkeyword.com. There is no result retrieved from searching “china day independence” and there is no impact on the rest search results.
• treason (rank: 2.2000)
The most of the search results are accessible. 5% to 25% of the search results have been blocked.

• BBC news (rank: 2.0909)
The most of the search results are accessible, but while searching some keywords, 5% to 40% of the search results have been blocked. The same situation occurs as searching “voice of America”. Therefore, the Chinese Internet control has the impact on foreign media.

• human rights (rank: 1.9000)
The most of the search results are accessible. 10% to 30% of the search results have been blocked.

• humanity (rank: 1.8000)
The most of the search results are accessible and are less impacted by Internet control. 10% to 40% among the 2 sets of search results have been blocked.

Internet control has had very limited or no impact on the following keywords:
• coup (rank: 1.0000)
• disciple (rank: 1.0000)
• oppression (rank: 1.0000)
• overthrow (rank: 1.0000)
As there is no sub keyword retrieved from Enteryourkeyword.com, the above keywords has been searched, the keywords do not contain any relation specific to China, so there is very limited or no impact on these keywords by Chinese Internet control.
Factors

From the research results, we can clearly find that some of the search results have been blocked or controlled in China. Analyzing and grouping this sensitive content and censored keywords into different topics or areas allowed an answer to the question “What are the factors that influence control of Internet search engines in China?”. Therefore, different factors which are influencing the control of Internet search engines in China could be defined according to the above list.

The research keywords were chosen according to different areas: political/religious, cultural/moral, and economic. Therefore, the factors which are influencing the control of Internet search engines in China found through this research were also grouped into these areas, and these factors were associated with different keywords in the keyword list which have been used in this research. The following table shows the keyword, rank, category/area and factor:

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Rank</th>
<th>Category/Area</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>falun gong</td>
<td>4.0000</td>
<td>religious</td>
<td>Illegal religious movements</td>
</tr>
<tr>
<td>Sequence search of “falun”, “lungong”, “lunda”, “dafa”</td>
<td>4.0000</td>
<td>religious</td>
<td>Illegal religious movements</td>
</tr>
<tr>
<td>massacre china</td>
<td>4.0000</td>
<td>political</td>
<td>Independence movements</td>
</tr>
<tr>
<td>minghui</td>
<td>4.0000</td>
<td>religious</td>
<td>Illegal religious movements</td>
</tr>
<tr>
<td>sex china</td>
<td>3.5455</td>
<td>cultural/moral</td>
<td>Sex and homosexuality</td>
</tr>
<tr>
<td>playboy</td>
<td>3.5000</td>
<td>cultural/moral</td>
<td>Sex and homosexuality</td>
</tr>
<tr>
<td>anti communist</td>
<td>3.4000</td>
<td>political</td>
<td>Anti communist</td>
</tr>
<tr>
<td>aids china</td>
<td>3.3333</td>
<td>cultural/moral</td>
<td>Health</td>
</tr>
<tr>
<td>Keyword</td>
<td>Rank</td>
<td>Category/area</td>
<td>Factor</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------</td>
<td>-------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>protect rights</td>
<td>3.3000</td>
<td>economic</td>
<td>Economic</td>
</tr>
<tr>
<td>dictatorship</td>
<td>3.1818</td>
<td>political</td>
<td>Anti communist</td>
</tr>
<tr>
<td>june 4</td>
<td>3.0833</td>
<td>political</td>
<td>Anti communist</td>
</tr>
<tr>
<td>genocide china</td>
<td>3.0000</td>
<td>political</td>
<td>Independence movements</td>
</tr>
<tr>
<td>hongzhi</td>
<td>3.0000</td>
<td>religious</td>
<td>Illegal religious movements</td>
</tr>
<tr>
<td>democracy china</td>
<td>2.8889</td>
<td>political</td>
<td>Opposition parties</td>
</tr>
<tr>
<td>voice of America</td>
<td>2.8000</td>
<td>political</td>
<td>News</td>
</tr>
<tr>
<td>tvbs</td>
<td>2.6000</td>
<td>political</td>
<td>News</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Political processes in Taiwan</td>
</tr>
<tr>
<td>dalai lama</td>
<td>2.3636</td>
<td>religious</td>
<td>Independence movements</td>
</tr>
<tr>
<td>independence china</td>
<td>2.3333</td>
<td>political</td>
<td>Independence movements</td>
</tr>
<tr>
<td>treason</td>
<td>2.2000</td>
<td>political</td>
<td>Anti communist</td>
</tr>
<tr>
<td>BBC news</td>
<td>2.0909</td>
<td>political</td>
<td>News</td>
</tr>
<tr>
<td>human rights</td>
<td>1.9000</td>
<td>political</td>
<td>Anti communist</td>
</tr>
<tr>
<td>humanity</td>
<td>1.8000</td>
<td>cultural/moral</td>
<td>Violence</td>
</tr>
<tr>
<td>coup</td>
<td>1.0000</td>
<td>political</td>
<td>Anti communist</td>
</tr>
<tr>
<td>disciple</td>
<td>1.0000</td>
<td>religious</td>
<td>Illegal religious movements</td>
</tr>
<tr>
<td>oppression</td>
<td>1.0000</td>
<td>political</td>
<td>Independence movements</td>
</tr>
<tr>
<td>overthrow</td>
<td>1.0000</td>
<td>political</td>
<td>Anti communist</td>
</tr>
</tbody>
</table>

Table 2 – Keywords, ranks, category/area and factor
Based on the results presented in Table 2 above, main themes can be further grouped together. This is shown in Table 3 below and represents the primary factors of internet control and topics that fall within them.

<table>
<thead>
<tr>
<th>Political/religious factors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>News</strong></td>
</tr>
<tr>
<td>-- Such as any news which critic on Chinese political/religious process and human rights issues…, refer to keyword list: TVBS; BBC news; Voice of America.</td>
</tr>
<tr>
<td><strong>Blogs</strong></td>
</tr>
<tr>
<td>-- Such as any blog which critic on Chinese political/religious process and human rights issues…</td>
</tr>
<tr>
<td><strong>BBS forums</strong></td>
</tr>
<tr>
<td>-- Such as any BBS forums which critic on Chinese political/religious process and human rights issues…</td>
</tr>
<tr>
<td><strong>Opposition parties</strong></td>
</tr>
<tr>
<td>-- Such as Taiwan political parties…, refer to keyword list: democracy China.</td>
</tr>
<tr>
<td><strong>Political processes in Taiwan</strong></td>
</tr>
<tr>
<td>-- Refer to keyword list: TVBS.</td>
</tr>
<tr>
<td><strong>Anti communist</strong></td>
</tr>
<tr>
<td>-- Such as Tiananmen June 4 movements…, refer to keyword list: anti communist; dictatorship; June 4; treason; human rights; coup; overthrow.</td>
</tr>
<tr>
<td><strong>Independence movements</strong></td>
</tr>
<tr>
<td>-- Such as Tibet and Xinjiang independence movements…, refer to keyword list: independence China; massacre China; genocide China; Dalai Lama; oppression.</td>
</tr>
<tr>
<td><strong>Illegal religious movements</strong></td>
</tr>
<tr>
<td>-- Such as Falun Gong group…, refer to keyword list: Falun Gong; sequence search of “falun”, “lungong”, “lunda”, “dafa”; minghui; Hongzhi; disciple.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cultural/moral factors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex and homosexuality</strong></td>
</tr>
<tr>
<td>-- Refer to keyword list: playboy; sex China.</td>
</tr>
</tbody>
</table>
Health
-- Such as Aids (Acquired Immure Deficiency Syndrome), STDs (Sexually Transmitted Disease) …, refer to keyword list: aids china.

Violence
-- Refer to keyword list: humanity.

**Economic factors**
-- Refer to keyword list: protect rights.

<table>
<thead>
<tr>
<th>Table 3 – Factors and sub categories/areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>These factors and sub topics will be discussed and analyzed in the next chapter “Findings analysis” below.</td>
</tr>
</tbody>
</table>

**Findings analysis**

According to the data from this research which have been summarized in the tables at the end of the previous sub chapter, we can now gain a view and understanding of what the primary factors and issues that affect internet control in China are. This sub chapter summarizes these.

Contents relate to Falun gong have attracted huge attention from Chinese Internet control system. Therefore, other content related similar religious groups are likely to receive the same attention from the Chinese Internet control system. The content related to Tibet and Xinjiang are also sensitive. This is not only a religious factor but also political factor as well. On the other hand, the contents relate to independence movement and anti communist are also sensitive. Therefore, these political and religious factors have taken the top spots on the Chinese Internet censor list. What contents has been controlled and why, has been covered earlier. It is clear that illegal religious movements, independence movements and anti communism factors are truly controlled by the Chinese government. Also according to the data from this research, content related to opposition parties and political progresses in Taiwan has also attracted attention from the Chinese Internet
control system. Therefore, other content related to dissident/democracy factors will also likely be impacted by the Chinese Internet control system. Besides these factors, news and other Medias such as blogs and forums which often broadcast and discuss these factors have brought attentions to the Chinese Internet control system as well. Therefore, these news, blogs, forums have become other factors which will be monitored and controlled. This is expanded to dissident/democracy websites, wiki/education websites, news websites, government websites, Taiwanese websites are all controlled in China. Compared to what content has been controlled and why it has been controlled, in the previous literature review chapter, it is clear that opposition parties, political progresses in Taiwan, news, blogs and forums factors are truly controlled by the Chinese government. As the world’s largest communist country pursuing a one party political policy, China has been attacked by many other western countries. In response, China has tightened its political control on all aspects. Therefore, it is not a surprise at all that the Chinese Internet control system has paid huge attention to all these factors.

According to the data from this research, content related to adult magazine (playboy) and sex china have brought huge attention from the Chinese Internet control system. Therefore, other content related to sexual materials very likely will receive the same attention. This is expanded to all sexually and homosexually explicit content and other content which may influence the Chinese culture ethics. According to the data from this research, content related to humility have been watched by the Chinese Internet control system. This is expanded to all violence factors which may influence human moral issues. Compare this to what content has been controlled and why they have been controlled in the previous literature review chapter, this appears to have the same reflection from all culture groups.

This research has not involved many economic keywords to be test with. It is because with the frequent update with search engines’ database, the economic factors can be only tracked when economic conflict happens. During the time of gathering data and selecting keyword, there is no such economic factor taking place. But this research still tested
keyword “protect rights”. It is considered as a controlled factor against Chinese Internet control system.

As this research stated at the beginning, in order to enter and stay in the Chinese Internet industry and market, both Chinese domestic and foreign Internet search engine companies have to follow Chinese regulations, rules and laws. So it is important that these companies understand these regulations, rules and laws, and most important those factors which have not been stated within them. This research has identified keywords and how they and the topics they are related to be dealt with by the search engine filters in China. Once Internet search engine companies understand these and avoid putting such information and contents within their database, they will avoid undesirable problems. So when the Internet search engine companies want to put any content and keyword link related to the above factors into their online search databases, the Internet search engines have to be aware of the Chinese Internet control issues.

**Limitations of the research**

By producing the analyzed keywords Excel spreadsheet, there are some facts and limitations during implementing this research.

- **Limitation of choosing keywords.**
  
  When producing the list of searching keywords under each set of keyword, I used the keywords generation function on the Enteryourkeyword.com. ([http://www.enteryourkeyword.com/index.php](http://www.enteryourkeyword.com/index.php)) It produced a list of related keywords and phrases, but some of them were not related to China, for example, when I entered “dictatorship”, it produced many phrases about Cuba and Argentina dictatorship. It would be better to have more phrases about China dictatorship, it would have given a different result while searching.

- **Page does not exist.**
  
  The cause of green results are mostly because of the bad links, that means these links are not blocked by reason of Internet control, but rather are due to the server problem or the page does not exist anymore.
Evading Internet control in China

As usually there is nothing on this world that is perfect, so it is with the Internet control in China. Although it is the most complex control system around the world, it still has its backyard. And it leaves the possibility for the Internet users in China to evade the Internet control in China by getting around the blocks of the websites and firewalls, therefore, the Internet users can still get access much more freely to ideas and information.

The way of evading Internet control in China (from user’s aspect):

- Using proxy servers. The proxy server can be setup easily to be associated with the user’s own browser and it will allow the user to access blocked websites. Chinese Internet control also tries to prevent this aspect, to stop the use of proxy servers, but most of these proxy servers are still running and could provide effectively services.
- Exploring the new Internet technologies to access more information. Such new technologies may include: Internet phone, BBS, blog, email, P2P, messenger etc.
Chapter 6: Conclusions

The latest news from Sydney Morning Herald on 6th of January 2008 “China to tighten internet control” (Sydney Morning Herald, 2008) reports that China has released their new regulations to control the popular video sharing websites. This is just an example showing that China intends to keep adding Internet control regulations to their existing systems. In China the Internet will not be able to access materials and contents containing sex, violence or gambling as well as anything against state interests or security or that threatens social morals. The latest online content related to independence movements in Tibet and Xinjiang have also been tightly control by the Chinese government. Besides the Chinese government, both large domestic and international search engines such as Google and Baidu are applying their own self-censor system in China to block certain sensitive contents. They do not and will not display any information and content that is critical of China in order to stay in the Chinese Internet market.

So what has happened to Internet search engines in China? After conducting this research, aspects which impact on Internet search engines operating in China have been documented. With the world’s largest population, China certainly is the most important market for both domestic and international companies. Each successful global company is willing to enter the Chinese market and would like to earn great benefit from it. This is the beginning of the Google case in China. When the Google users found that they had struggled with Google.com services in China, Google tried to solve such problem by launching a Chinese domestic presence “Google.cn”. But as Google.cn is based in China, Google has to follow the laws, rules and regulations of China. The final result is Google had a deal with Chinese government and Google.cn was launched in February 2006.

The literature review was carried out at the beginning of this research, and a list of censored keywords was chosen. Then this research involved simultaneous use of these keywords in both China and New Zealand in order to find out what sensitive content and censored keywords which are impacted by the Chinese Internet control. Sensitive content has been removed from the search results of Internet search engines in China. The
analysis of the research data and information answers the research question “what are the factors which influence the control of Internet search engines in China?”.

To conclude this research, explicit answers to each research question follows.

- **What are the factors which influence the control of Internet search engines in China?**
  
  The problem of this research was defined and main research question was asked “what are the factors which influence the control of Internet search engines in China?” and some answers for it have been found. The factors were grouped and categorized into 3 different areas. They are political/religious factors which include news, blogs, BBS forums, opposition parties, political process in Taiwan, anti communist, independence movements, illegal religious movements; cultural/moral factors which include sex and homosexuality, health, violence; and economic factors. Among these factors, Chinese government has blocked most political/religious and cultural/moral factors related information and contents. 17 out 22 keywords whose rank is higher than 1.0000 belong to political/religious factor. From this research, the illegal religious movement factor has been found as the highest rank to be blocked, followed by sex and homosexuality factor, independence movements factor and anti communism factor. They are the top 4 highest rank factors which have been controlled by the Chinese Internet control system.

- **What are the current factors/issues for Internet development and control in China?**
  
  “The use of information technology is vital for the world economy and social development” stated by former Chinese Prime Minister Zhu Rongji, at a technology lecture in Beijing on 27 February 2002. (Amnesty International, 2002)

  In fact, this statement also applies in China. The Internet and Internet search engines as the information technologies of the new millennium, have been brought closer and closer to a Chinese citizens’ everyday life. Internet and Internet search engines have changed the world and will bring more and more challenges to everyone all over the world. China currently has over 137 million
Control of Internet search engines in China

Internet users and in the coming one or two years China will soon replace the USA as the largest Internet market around the world. Chinese citizens could access any information they want by sitting in front of a computer and searching on search engines. Although China is still controlling the Internet for some reasons, but compare to the time before year 1995 that Internet has been brought into China, the Internet has become a force for changes in China. This research examined the current situation of Internet control in China and showed that China is developing its own Internet border.

- What are the current factors/issues for Internet control in other countries, such as USA and Vietnam?

In this research, the Internet control issues of other countries such as USA and Vietnam have also been explored. So what are the current factors/issues for Internet control in USA and Vietnam? As this research have been conducted based on New Zealand and China, USA is considered as having less Internet control issues like New Zealand. The research results and decision matrix in the Excel spreadsheet shows that while the total number of search results and number of working links retrieved in New Zealand $\geq$ the total number of search results and number of working links retrieved in China, the certain keywords has less or no impact by Internet control issues. Most of these keywords are in the unmarked or green categories. USA only controls the contents which contain moral factors such as violence contents. Vietnam has the similar situation as China. The Vietnam government has tried to setup their laws, rules and regulations for their Internet, therefore, the Vietnam government could block any undesirable information like Chinese government does.

Overall China has an effective system which has been used to control its Internet so far. There is information and content which has been controlled and other content which has not. From the data over time, it clearly shows that the Chinese Internet control system maintains a watching list and keeps updating it actively. So will China keep controlling the Internet and will Chinese government be successful on controlling the Internet? In my
opinion, it will be much more difficult to control the Internet in the future, so it will be very interesting to see whether China will lose control of its Internet or whether China will invent a more accurate system to control its Internet and be successful in continuing its Internet control.

**Further research**

The Internet has been controlled efficiently in China so far. So how does this take place? To answer this question, people will easily answer that because of the Chinese government's dictatorship and powerful control of the Internet. But there are still other opinions, the differences between the values of people’s thought in each country would also affect the Internet control result. For example, in poor countries and most developing countries, people do not care about the rights of freedom of speech as much rights as people in developed countries. They may think that the freedom of speech is just a luxury idea for them. So in future research, the researchers should also research on the psychology of people when they are controlled when using the Internet.
References


Control of Internet search engines in China


Appendix

Appendix 1 – Keywords list

The following keywords list was conducted by using the keyword generator tool on the Enteryourkeyword.com website.

“democracy china”
American china crisis democracy entrepreneurship future in series shadow
china democracy
china democracy party
china chinese democracy first in life political republic taiwan
build china cooperative democracy industry story
china deliberative democracy in search
authoritarianism china conflict democracy ideological in modern
china democracy in republican socialism
protest province report university

“human rights”
human right
human right watch
to protect human right worldwide
human right campaign
human right violation
universal declaration of human right
international human right
human right commission
human protecting right worldwide
china human right
“dalai lama”

dalai lama quote
dalai lama and mit
corposition dalai lama story tibet
destructive emotion a scientific dialogue with the dalai lama
autobiography dalai his holiness lama land original people tibet	
tibet dalai lama
dalai lama biography
dalai essential his important lama teachings
dalai his holiness lama living wisdom
dalai ethics his holiness lama millennium new
dalai lama book

“BBC news”

bbc news
bbc world news
bbc world latest news
bbc arabic news
bbc news uk
bbc news africa
bbc urdu news
bbc world news news
bbc execution hussein news saddam tape
bbc sports news
bbc news 24
“tvbs”
  tvbs news
  tvbs taiwan
  tvbs asia
  tvbs tw
  tvbs news taiwan

“voice of America”
  voice of america radio
  america culture deaf from in voice
  voice of america voa
  america black girl in voice young
  voice of america persian
  voice of america amharic
  america page radio voice
  america bangla news radio voice
  voice of america special english
  voice to america radio show

“falun gong”
  china edition falun gong revised
  challenge china falun gong
  china falun gong
  falun gong cult
  enlightenment excercises falun gong health perfect principle
  day end falun gong
  complete falun falun gong teachings zhuan
  deceiving falun gong his hongzhi li life public ruining
  ethnographic falun gong in state study united

Sequence search of “falun”, “lungong”, “lunda”, “dafa”
“hongzhi”
   li hongzhi
decieving falong gong his hongzhi li life public running

“minghui”

“disciple”

“humanity”
   against crime humanity war war world
   against crime definition humanity
   against crime humanity record
   alabama habitat humanity
   alliance humanity new
   angeles habitat humanity los
   antonio habitat humanity in san
   antonio habitat humanity san
   application habitat humanity
   architecture humanity

“treason”
   cold from liberal terrorism treachery treason war war
   mortal treason
   cue for treason
   act of treason
   none dare call it treason
   definition of treason
   act high treason war
   treason the new world order
   cue for treason summary
   high treason
“dictatorship”
  advantage dictatorship
  advantage dictatorship disadvantage
  argentina dictatorship in
  argentina dictatorship military
  cause dictatorship
  country dictatorship
  cuba dictatorship
  cuba dictatorship in
  definition dictatorship
  democracy and dictatorship
  democracy dictatorship monarchy

“june 4”
  2005 4 brian hepp june karen sullivan wedding
  2006 25 4 day june sunday
  4 7 day israel june
  2006 2006 4 hawaiian honolulu june oahu vacation
  june 4 2003
  12 4 day june
  12 4 day june part2
  june 4 1989
  1 4 day israel june
  1988 4 june our wedding
  june 4 china
  4 june ride

“oppression”

“overthrow”
“coup”

“protect rights”
  to protect human right worldwide
  protect your right
  either fully protect right thats
  protect human right
  how to protect resident right
  protect your digital right
  employment protect right
  from protect right yourself
  home nursing protect resident right
  either fully protect right

“independence china”
  china independence
  china independence ironstone
  china day independence

“anti communist”
  american anti communist death expose have infiltrated inside latin league league
  nazi shocking squad terrorist world
  anti communist europe hitler in sentiment
  a covert life jay lovesone communist anti communist and spymaster
  anti communist crusade eisenhower
  anti communist mccarthyism

“massacre china”
  china jiangsu massacre nanjing nanking sheng
“genocide china”
genocide in china

“playboy”
.com magazine playboy
.com playboy
.com playboy store
.com playboy tv
1 centerfold page playboy
10 big girl playboy
10 girl pac playboy
1024 playboy wallpaper
12 big girl photo playboy
12 big girl playboy

“sex china”
china sex
china sex tape
china sex com
china sex video
china sex movie
sex teen china
free china sex
china woman sex
china sex picture
china sex photo
china doll sex tape
“aids china”
aids in china
aids china cultural eating in politics rice southwest spring
hiv and aids in china
Appendix 2 – Research results
Appendix 3 – Decision matrix