# Hotel Information Exposure in Cyberspace: The Case of Hong Kong

Rosanna Leung, and
Rob Law

[a]School of Hotel and Tourism Management
The Hong Kong Polytechnic University, Hong Kong
rosanna.leung@connect.polyu.hk
hmroblaw@polyu.edu.hk

## Abstract

Search engines are an everyday tool for Internet surfing. They are also a critical factor that affects e-business performance. This study compares the quantity of hotel web pages and pages indexed by search engines. Values of index ratio indicated that international chain hotels were not well indexed by domestic search engines. Furthermore, the web servers of local chain and independent hotels of Hong Kong contain a large amount of outdated or unlinked materials. Though how search engine index pages cannot be controlled, hotel managers could control what should be indexed. Web masters were recommended to relocate these materials to independent folders and use robots exclusion protocol to restrict the access of web crawlers and robots.

**Keywords:** index ratio, robots exclusion protocol, search engine, web content, hotel website

## 1 Introduction

No matter how rich the information provided on a website (functionality) is and how user-friendly the website is (usability), if a website cannot be found from a search engine, it remains unknown to potential customers and business partners. Web surfing currently becomes a major information search channel. As such, tourism organizations have adopted the Internet as one of the primary marketing channels for attracting customers (Buhalis & Law, 2008). Nowadays, many companies have spent plenty of resources to create and maintain their websites. However, a large amount of investments on the web does not necessarily mean more business in return (Hernández, Jiménez, & Martín, 2009). Most Internet users could only memorize few popular web addresses (Uniform Resource Locator, URL). They simply search their desired information by typing the keywords and locate related result from search engine results. In 2001, 28 million websites were available; whereas a decade later, it has ten-fold increase to 298 million websites (Netcraft, 2011).

Apparently, the number of websites is growing in million sites each month. Researchers found that over 80% of web surfers use search engines to locate their desired information (Fox, 2002, Lawrence & Giles, 2000), and 78% of Internet users search online product information before they buy (Pew, 2011). There has been evidence indicating an increasing number of tourism enterprises using the Internet as a marketing tool, and more travel planning being conducted via the Internet (Buhalis & Law, 2008, Pan & Fesenmaier, 2006). As accommodation is one of the most important components for travellers, hotel websites will be a main search target for travellers, and for this reason, hotel websites are examined in this study.

Currently, the most popular international search engines were Google, Yahoo, Bing, Ask, and AOL (Hitwise, 2011). Making a website visible to search engines is one of the critical success factors on hotel marketing. However, most hotels only focus on maintaining their websites' content up-to-date and they are not aware of the visibility and freshness of the information of their websites on search engines. Furthermore, the indexing power of search engine crawlers and robots was enhanced significantly in the past decade. As such, many web contents incorporated Flash, and PDF can now be searched by visitors via search engines. However, web masters may not want search engine spiders to collect the whole web server's content. Therefore they should setup instructions that bound spiders fetching the specific web contents.

In view of the arising challenge, this research study is:

- to analyze the web information richness in search engine databases of various hotel categories;
- to compare the crawling result of the top global and domestic search engines; and
- to examine the search engine index ratio among hotel categories.

## 2 Research Background

### 2.1 Information on the Internet

The Internet is one of the most cost effective and efficient information and marketing channels that has no geographical and time barriers (Kasavana, Knutson, & Polonowski, 1997, Schmidt, Cantallops, & dos Santos, 2008, Tjostheim, Tussyadiah, & Hoem, 2007). According to Peterson, Balasubramanian, and Bronnenberg (1997), the Internet provides the capability of inexpensively storing a vast amount of information in different virtual locations. With the growing importance of search engines, the information quality that search engines provide should match customers' need. Prior studies have confirmed good quality of information provided by search engines must be up-to-date, available and visible, complete, popular, and with an adequate amount of information (Klein, 2002, Knight & Burn, 2005, Zhu & Gauch, 2000). Due to the huge volume of web pages being updated on each day, search engines delete the removed page information in around 10 and 26 days for Google and Yahoo (Lewandowski, 2008). However, if a page is out-dated but remains on the web server, search engines will retain the records in their databases. Furthermore, the popularity could affect influence the search engine ranking (Alexa, 2011).To increase the popularity of a website, more external links pointing to the URL are always associated with a larger amount of drive in traffic. Prior studies have indicated the importance of search engine marketing (Paraskevas, Katsogridakis, Law, & Buhalis, 2011, Xiang & Pan, 2009). However, the prerequisite of search engine marketing is that web contents must be stored in search engine databases; therefore the content indexing should not be overlooked.

### 2.2 Search Engine Indexing

Search engines collect web contents by sending out crawlers and robots (commonly known as spiders), which are the software agents that search engines employed to collect content for their databases (Kobayashi & Takeda, 2000). The basic web components for displaying a web page on the web browser included Hypertext

Markup Language (HTML) / Extended Markup Language (XML) codes, photos, animated graphics, and videos. In addition, to better control the web content, web programming codes such as JavaScript, Flash, PHP, ASP, CGI are included (Lecky-Thompson, 2008). With the increasing amount of broadband penetration, web surfers can enjoy high quality video and audio entertainments. In order to create more interactive and animated websites, Flash becomes one of the most popular web development tools. In 2001, many web contents generated by Flash, ZIP, and PDF were unable to be indexed by search engines. As such, it was not recommended that web designers to use these tools for web design (Sherman & Price, 2001). With the advancement of the search engines' crawling power, most spiders are now able to retrieve information from these file formats so web designers could have more choices to make attractive websites without worrying about search engine indexing and ranking issues.

To ensure a hotel website is listed by major search engines, a simple way is to register in a search engine's directory. All search engines encourage web masters to register their URLs onto their databases so their spiders can visit and index the content (Baidu, 2011; Bing, 2011a; Google, 2011a; Yahoo, 2011a). No search engine guarantees all pages are indexed, and they limit crawling frequency and the depth so the new websites will remain "invisible" until the next update crawl (Wouters, Reddy, & Aguillo, 2006). To increase crawling depth and indexing coverage, search engine engineers suggest web masters to submit sitemaps (Bing, 2011b, Google, 2011b), and enhancing web design style (Google, 2011b, Yahoo, 2011b). This is to ensure that their spiders could collect more accurate information and ensure all pages could be crawled without any missing pages. However, not all the pages inside a web server should be visible to visitors. As a result, the robot exclusion protocol was introduced in 1994 for controlling the access right of web crawlers and robots (Zittrain, 2008). Findings of a prior study discovered that 8% of the web content crawled by the search engine spiders were outdated or inactive (Al-Masri & Mahmoud, 2008) but the study did not include those contents that should not be visible to visitors such as system messages and software application program files. This hints some web masters did not manage their website contents well. Therefore, this study is to examine the current situation of content management of Hong Kong hotel web masters.

## 3 Research Methodology

This study examined the index ratio of Hong Kong hotel websites by counting the number of hotel pages and the pages indexed by search engines. The index ratio will be further explained in Section 4.3. In total, there were 150 hotels listed on Hong Kong Tourism Board's website (HKTB, 2011) and their names and URLs were retrieved. Two sets of data were collected in July 2011 from the Internet. The first batch of data was collected via sitemap generation software. By providing the hotel URL, the software scanned the whole tree structure from the home page within the domain and generated the sitemap of the hotel website. Because the programming languages adopted by each hotel website were different (e.g. HTML, PHP, ASP, and Flash), and the crawling methodologies of the sitemap generation software were different, this study adopted two sitemap generation applications. The sitemaps created by both sitemap generators were manually reviewed by the authors. After that,

the number of information pages and number of languages available were recorded into the database. Several chain hotel websites shared the information among hotels. So it is impossible to separate the pages by individual hotels. Therefore, these websites were excluded in this study, resulting in 127 hotel websites included for data analysis. The second batch of data was collected from search engines. According to Hitwise (2011), the top three search engines by search volume were Google (65.45%), Yahoo (15.66%) and Bing (12.97%). With around 60% of the tourists visiting Hong Kong were from mainland China (hereafter known as China), the website ranking in China's search engines could thus affect a hotel's business performance. In China, the leading search engine Baidu ranked six in Alexa (2011) which took up 65% of the search volume. As a result, these four search engines were included in this study. The command "site:*domain name*" was used to list out the pages crawled by the search engines. The total numbers of pages of each search engine were also recorded in the database.

## 4 Findings and Discussions

The hotel websites analyzed in this study were classified into four categories. An international chain refers to a hotel chain which has sister hotels in multiple continents. An Asia-based hotel chain refers to the chain that manages hotels within Asia. A local chain relates to the hotel network in Hong Kong, and independent refers to those that are not affiliated with any hotel networks. In this study, out of 127 hotels, 22 were affiliated with 10 international chains, 15 were affiliated with six Asia-based chains, 48 were part of 17 local chains, and 42 hotels were operated independently.

### 4.1 Content Pages on the Hotel Websites

In this study, a majority of the hotel websites provided content with more than one language, but six websites only had English pages. Among these six hotels, four of them were independent hotels and two of them were international chain hotels. Many chain hotels had more than one website. One of the websites was hosted by the chain, in which the domain name is the name of the chain. The other website's domain is the name of the hotel, which is normally owned by the hotel and maintained by hotel staff. Table 1 illustrates the average number of web pages in each category. The result indicated that the international chain, on average, offered 6.45 languages which were significantly more than the other three categories. For local chain and independent hotels, they normally offered English and Chinese (traditional and/or simplified) versions only. Asia-based chain websites additionally offered Japanese and Korean pages. For international chain hotels, they further provided pages in Spanish, French, German, and/or Portugal.

When the total number of web pages was examined, international chain hotels provided 370 pages and Asia-based chain hotels offered 380 pages. Both of them were two and three times more than local chain and independent hotels. Apparently, the number of available languages directly affects the total number of web pages. When the total number of pages on a hotel website is divided by number of available languages, the result shows a different view. The average number of pages per language for international chain hotels was 64 pages and the corresponding number for Asia-based chain hotels was 95 pages. Local chain hotels and independent hotels

were quite similar with 39 and 33 pages per language. As a result, the amount of web content information that Asia-based chain hotels provided was three times richer than local chain and independent hotels, and they were 30% richer than international chain hotels.

**Table 1.** Number of Hotel Web Pages by Hotel Categories

| | International Chain | Asia-based Chain | Local Chain | Independent Hotel | Total | $\chi^2$ | Sig. |
|---|---|---|---|---|---|---|---|
| N | 22 (17.32%) | 15 (11.81%) | 48 (37.80%) | 42 (33.07%) | 127 (100%) | 17.18 | .001* |
| No. of chains | 10 | 6 | 17 | - | | | |
| *No. of Languages Provided* | | | | | | F | Sig. |
| Mean | 6.45[1,2,3] | 4.20[1] | 3.38[2] | 2.68[3] | 3.80 | 18.31 | .000* |
| Std. | 3.83 | 1.86 | 1.34 | 0.93 | 2.37 | | |
| *Total No. of Pages* | | | | | | | |
| Mean | 369.05[4,5] | 380.13[6,7] | 148.76[4,6] | 93.59[5,7] | 197.98 | 8.19 | .000* |
| Std. | 533.61 | 242.86 | 147.45 | 107.39 | 285.16 | | |
| *Average No. of Pages per Language* | | | | | | | |
| Mean | 64.30 | 94.87[8,9] | 39.60[8] | 32.64[9] | 48.44 | 8.06 | .000* |
| Std. | 81.08 | 42.76 | 30.86 | 34.89 | 50.07 | | |

[1, 4, 6] *The mean difference for these two sectors is significant at the 0.05 level: p = 0.005, 0.010, and 0.020*

[2, 3, 5, 7-9] *The mean difference for these two sectors is significant at the 0.05 level: p = 0.000*

* *Significant at p<0.05*

### 4.2 Pages Indexed by Search Engines

The crawling power of Google's spider seems stronger than the other three. In this study, among all hotel categories, Google indexed almost twice the number of pages than other search engines. Asia-based chain ranked the first with 288 pages indexed, local chain ranked second with 192 pages, and international chain and independent hotels had 171 and 151 pages crawled by Google's spider. The indexing power for both Yahoo and Bing were quite similar (Table 2). These three search engines indexed all the hotels listed on HKTB's website. The domestic search engine, Baidu, indexed the least number of hotel web pages and was unable to index all listed hotels. Three local chain and six independent hotel websites were not found in Baidu's database. For international chain hotels, Baidu listed out all the web pages by the domain name.

However, it was unable to filter out the pages that belong to designated hotels. Therefore, the total number of pages was unable to retrieve these four hotels. Among all search engines, the numbers of pages indexed for Asia-based chain and local chain were higher than international chain hotels. For Baidu, the total number of pages crawled for all hotel categories was 30% less than the Google's record. Particularly for international hotels, Baidu only indexed 15% of Google's contents.

**Table 2.** Number of Pages Crawled by Search Engines by Categories

| | Google | Yahoo | Bing | Baidu | Total | F | Sig. |
|---|---|---|---|---|---|---|---|
| *International Chain* | | | | | | | |
| N | 22 | 22 | 22 | 17 | | 10.396 | .000* |
| Mean | 171.05[1,2,3] | 90.14[1] | 65.19[2] | 26.07[3] | 93.83 | | |
| Std. | 137.29 | 64.09 | 44.68 | 24.20 | 97.29 | | |
| Min | 33 | 25 | 5 | 0 | | | |
| Max | 503 | 331 | 156 | 75 | | | |
| *Asia-based Chain* | | | | | | | |
| N | 15 | 15 | 15 | 15 | | 17.391 | .000* |
| Mean | 288.53[4,5,6] | 138.73[4] | 153.80[5] | 90.64[6] | 169.24 | | |
| Std. | 97.19 | 81.04 | 81.28 | 37.23 | 106.04 | | |
| Min | 12 | 8 | 8 | 5 | | | |
| Max | 395 | 835 | 353 | 149 | | | |
| *Local Chain* | | | | | | | |
| N | 48 | 48 | 48 | 45 | | 12.900 | .000* |
| Mean | 191.78[7,8,9] | 80.44[7] | 81.42[8] | 38.40[9] | 99.02 | | |
| Std. | 212.99 | 76.46 | 75.38 | 34.58 | 133.21 | | |
| Min | 12 | 2 | 2 | 1 | | | |
| Max | 897 | 327 | 324 | 158 | | | |
| *Independent Hotel* | | | | | | | |
| N | 42 | 42 | 42 | 37 | | 6.137 | .001* |
| Mean | 151.39[10,11,12] | 64.15[10] | 67.98[11] | 45.78[12] | 83.47 | | |
| Std. | 208.00 | 68.29 | 72.91 | 54.75 | 125.79 | | |
| Min | 1 | 1 | 1 | 1 | | | |
| Max | 867 | 362 | 361 | 223 | | | |

[1-9] *The mean difference for these two sectors is significant at the 0.05 level: p = 0.000*

[10-12] *The mean difference for these two sectors is significant at the 0.05 level: p = 0.007, 0.011, and 0.001*

* *Significant at p< 0.05*

### 4.3 Index Ratio

This study further examined the search engine index ratio. This ratio was calculated by the following formula:

$$\textit{Index Ratio} = \frac{\textit{Total number of pages indexed by a search engine}}{\textit{Total number of content pages on a hotel website}} \qquad (1)$$

When the ratio is less than one, that means a hotel website is not fully indexed by the search engine. If the ratio is greater than one, which implies the search engine over indexed some web contents such as outdated promotional materials or system messages which should be invisible to visitors. Fig 1 shows a contrasting index ratio pattern for Google and Baidu. There were a large number of the hotel websites (90%) which Google has indexed at least 50% of the web pages. However, for Baidu, less than 30% of the hotel websites has over 50% index ratio. For both Yahoo and Bing, their index ratios were quite similar with around 60% of the websites received at least 50% index ratio. However, the over index ratio for Google was also very high. Over 60% of the hotels received over 100% indexing, and 21% of them received over 200% indexing. For Yahoo and Bing, only 23% of hotel websites were over indexed. For Baidu, only 7% was over indexed. This hints that visitors could have higher opportunities to obtain dated information from Google search result or could easily access some system function pages which could cause system errors.

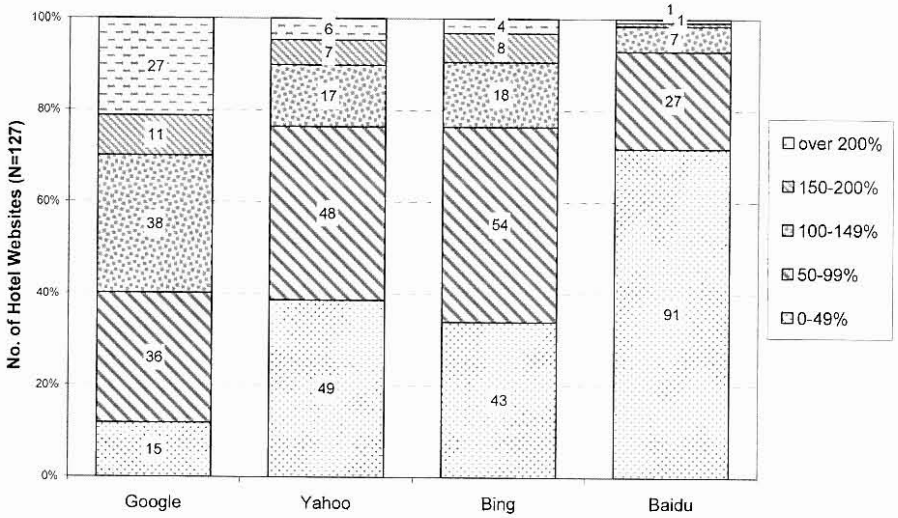


**Fig 1.** Frequency Diagram of Hotel Websites Search Engine Index Ratio

When the result was further divided into four hotel categories, more interesting findings were received. Although the average total number of web pages for independent hotels was only 94, Google has crawled an average of 208 pages. This gave independent hotels the highest index ratio with an average of 176% indexed by Google and 51% indexed by Baidu. In this study, Google even indexed one independent hotel with 10 times of its total number of web pages. The excess pages were mainly dated events information in 2010 and 235 pages of online reservation templates. Local chain hotels also have a high index ratio with 42% for Baidu to 170% for Google.

After examining the content of these sitemaps, several websites' application function or system message pages were indexed by Google, and plenty of the unlinked content was outdated and inactive web content was stored on the web server. If the web

master did not control the search engine crawling behavior, the search engine spiders could crawl the whole web server and list out unwanted information to visitors.

**Table 3.** Hotel Website Index Ratio by Categories

| | Google | Yahoo | Bing | Baidu | Total | F | Sig. |
|---|---|---|---|---|---|---|---|
| *International Chain* | | | | | | | |
| N | 22 | 22 | 22 | 17 | | 5.46 | .000* |
| Mean % | 100.43[1,2] | 66.10 | 54.04[1] | 25.34[2] | 64.73 | | |
| Std. | 72.65 | 61.65 | 48.48 | 28.70 | 61.93 | | |
| Min | 5.52 | 3.02 | 1.13 | 0.00 | | | |
| Max | 257.00 | 269.12 | 151.28 | 76.92 | | | |
| *Asia-based Chain* | | | | | | | |
| N | 15 | 15 | 15 | 15 | | 5.16 | .000* |
| Mean % | 92.70[3,4] | 45.37[3] | 50.63 | 31.24[4] | 55.39 | | |
| Std. | 63.49 | 37.74 | 38.33 | 31.35 | 49.24 | | |
| Min | 13.33 | 5.14 | 37.55 | 11.89 | | | |
| Max | 280.51 | 159.32 | 159.32 | 126.27 | | | |
| *Local Chain* | | | | | | | |
| N | 48 | 48 | 48 | 45 | | 20.29 | .000* |
| Mean % | 170.49[5,6,7] | 77.87[5] | 76.44[6] | 41.53[7] | 92.43 | | |
| Std. | 141.91 | 53.49 | 42.48 | 37.31 | 93.92 | | |
| Min | 14.83 | 6.51 | 3.21 | 0.66 | | | |
| Max | 715.63 | 287.50 | 177.45 | 136.17 | | | |
| *Independent Hotel* | | | | | | | |
| N | 42 | 42 | 42 | 37 | | 9.12 | .000* |
| Mean % | 175.91[8,9,10] | 87.84[8] | 90.65[9] | 50.74[10] | 102.87 | | |
| Std. | 188.27 | 68.12 | 68.02 | 51.05 | 118.31 | | |
| Min | 6.25 | 3.48 | 3.48 | 0.00 | | | |
| Max | 1070.37 | 324.00 | 324.00 | 275.31 | | | |

[5-7, 10] *The mean difference for these two sectors is significant at the 0.05 level: $p = 0.000$*

[1-4, 8, 9] *The mean difference for these two sectors is significant at the 0.05 level: $p = 0.047$, 0.001, 0.027, and 0.003*

**Significant at $p < 0.05$*

## 5 Conclusions

This study examined the Hong Kong hotel website information richness and the page index ratio among search engines. The results indicated Asia-based chain hotels provided richer content than international chain hotels. Several international chain hotels not only disseminate information via their chain websites but also via the domain under hotel names. The content in hotels' own domain is richer than the corresponding one on the chain's website but the number of available languages was far less than the chain websites. Two of them only provided English pages. The average number of languages available on international chain hotel websites was 6.45 languages. The other hotel categories mainly provided English, together with traditional and simplified Chinese. Google's spiders got the highest crawling ability which enables the majority of hotel web pages indexed and become visible to Google

visitors. The crawling power of Baidu was the weakest. On international chain hotel websites, the pages indexed was only 15% of that from Google's. Though the indexing and crawling power of Google was the highest among all search engines, around 60% of the hotel websites were over indexed. This could make visitors locate the dated information or accidentally access part of the system which might affect the database accuracy. When comparing the index ratio among hotel categories, the local chain and independent hotels got the highest index ratio. Such a high ratio could be caused by dated web pages resided on the web server and the application program files. If the web master of these two categories did not manage their web content properly, the dated content and program files could be crawled by search engines. Finally, international chain hotels performed best in link popularity. This could help the hotel websites improve search engine ranking and driving in business (Alexa, 2011).

Hotel web masters should not only focus on the website indexing of international search engines, the performance of domestic search engines is also important. For instance, over 60% of tourists to Hong Kong were from China and over 65% of Chinese use Baidu as their primary search engine. As such, web masters should spend more effort on improving the indexing and ranking position of websites on Baidu. To prevent search engines from over indexing, it is recommended that web masters should implement the robots exclusion protocol to control search engines.

In case web masters prefer to keep the dated content for future reference, they should relocate the pages to separate folders. To prevent web spiders from crawling these specific folders, they must place a text file call robot.txt in the root of the website hierarchy. This file provides instructions allowing or limiting the search engine spiders fetching specific folders' content. If this file does not exist, the spiders will assume there are no restrictions on spider crawling of the web server.

## 6 Limitations and Future Research

This study has several limitations. First, the website hierarchy of international chain hotel websites was huge, and for this reason, the sitemap generation software was unable to fetch the complete web hierarchy in each case. Second, some websites used server-side programming for information retrieval but the sitemap generation software did not function properly so that pages using ASP or ASPX might be omitted. Furthermore, this study follows the hotel URL listed on HKTB's website. However, several hotels, especially chain hotels, had more than one domain name. Some of the hotels listed the chain URL but some listed the hotel domain.

Therefore, the results were not standardized. Finally, the tree structure of the website information arrangement could affect the fetch result. Some websites use languages as the name of the sub-folder where some websites use program variables to customize language settings. This could affect the search engine spider's performance. As mentioned by the top two search engines, enhancing the web design style could affect the performance of indexing and ranking (Google, 2011b, Yahoo, 2011b). Therefore, future research can further examine the relationship between index ratio and matching of search engines' preferred design style.

## References


Alexa. (2011). *Alexa Top 500 Global Sites*. Retrieved August 26, 2011, from http://www.alexa.com/topsites

Al-Masri, E. & Mahmoud, Q. H. (2008). Investigating web services on the world wide web. In *Proceeding of the 17th International Conference on World Wide Web - WWW'08* (pp. 795-804). Presented at the 17th International Conference on World Wide Web, Beijing, China.

Baidu. (2011). *Baidu Search Help Center-Site Registration*. Retrieved August 29, 2011, from http://www.baidu.com/search/url_submit.html

Bing. (2011a). *URL Submission - Bing*. Retrieved August 28, 2011, from http://www.bing.com/webmaster/SubmitSitePage.aspx

Bing. (2011b). *Getting started checklist*. Retrieved August 29, 2011, from http://onlinehelp.microsoft.com/en-us/bing/hh204497.aspx

Buhalis, D. & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet-The state of eTourism research. *Tourism Management* 29(4): 609-623.

Fox, S. (2002). *Pew Internet & American Life Project. Search Engines*. Retrieved September 25, 2009, from http://www.pewinternet.org/Reports/2002/Search-Engines.aspx

Google. (2011a). *Add your URL to Google*. Retrieved August 29, 2011, from http://www.google.com/addurl/

Google. (2011b). *Search Engine Optimization (SEO) - Webmaster Tools Help*. Retrieved August 29, 2011, from http://www.google.com/support/webmasters/bin/answer.py?answer=35291

Hitwise. (2011). *Search Engine Analysis. Hitwise United States - the power of competitive intelligence*. Retrieved August 26, 2011, from http://www.hitwise.com/us/datacenter/main/dashboard-23984.html

HKTB. (2011). *Search Hong Kong Hotels - Hong Kong Tourism Board*. Retrieved August 26, 2011, from http://www.discoverhongkong.com/eng/jsp/hotel/search-result.jsp

Kasavana, M. L., Knutson, B. J. & Polonowski, S. J. (1997). Netlurking: The Future of Hospitality Internet Marketing. *Journal of Hospitality & Leisure Marketing 5*: 31–44.

Klein, B. (2002). When Do Users Detect Information Quality Problems on the World Wide Web? In *Proceedings of the 8th Americas Conference on Information Systems (AMCIS 2002)*. Dallas, Texas.

Knight, S. & Burn, J. (2005). Developing a Framework for Assessing Information Quality on the World Wide Web. *Informing Science 8*: 159-172.

Lawrence, S. & Giles, C. L. (2000). Accessibility of Information on the Web. *Intelligence* 11(1): 32-39.

Lewandowski, D. (2008). A three-year study on the freshness of web search engine databases. *Journal of Information Science* 34(6): 817 -831.

Murphy, H. C. & Kielgast, C. D. (2008). Do small and medium-sized hotels exploit search engine marketing? *International Journal of Contemporary Hospitality Management 20*(1): 90-97.

Netcraft. (2011). *March 2011 Web Server Survey | Netcraft*. Retrieved April 4, 2011, from http://news.netcraft.com/archives/2011/03/09/march-2011-web-server-survey.html

Pan, B. & Fesenmaier, D. R. (2006). Online Information Search: Vacation Planning Process. *Annals of Tourism Research* 33(3): 809-832.

Paraskevas, A., Katsogridakis, I., Law, R. & Buhalis, D. (2011). Search Engine Marketing: Transforming Search Engines into Hotel Distribution Channels. Cornell Hospitality Quarterly 52(2): 200 -208.

Peterson, R., Balasubramanian, S. & Bronnenberg, B. (1997). Exploring the Implications of the Internet for Consumer Marketing. *Journal of the Academy of Marketing Science 25*(4): 329-346.

Pew. (2011, May). *What Internet Users Do Online | Pew Research Center's Internet & American Life Project.* Retrieved August 27, 2011, from http://www.pewinternet.org/Trend-Data/Online-Activites-Total.aspx

Rosen, D. E. & Purinton, E. (2004). Website Design: Viewing the Web as a Cognitive Landscape. *Journal of Business Research 57*(7): 787-794.

Schmidt, S., Cantallops, A. S. & dos Santos, C. P. (2008). The Characteristics of Hotel Websites and Their Implications for Website Effectiveness. *International Journal of Hospitality Management 27*(4): 504-516.

Sherman, C. & Price, G. (2001). *The invisible Web: uncovering information sources search engines can't see.* Information Today, Inc.

Tjostheim, I., Tussyadiah, I. P. & Hoem, S. O. (2007). Combination of Information Sources in Travel Planning a Cross-national Study. In *Information and Communication Technologies in Tourism 2007.* Wien: Springer-Verlag.

Wouters, P., Reddy, C. & Aguillo, I. (2006). On the visibility of information on the Web: an exploratory experimental approach. *Research Evaluation 15*(2): 107-115.

Xiang, Z. & Pan, B. (2009). Travel Queries on Cities in the United States: Implications for Search Engine Marketing in Tourism. In *Information and Communication Technologies in Tourism 2009.* Amsterdam, The Netherlands: Spring-Verlag.

Yahoo. (2011a). *Yahoo! Directory Suggest a Site.* Retrieved August 29, 2011, from http://add.yahoo.com/fast/add

Yahoo. (2011b). *SEO basics | Yahoo! Style Guide.* Retrieved August 29, 2011, from http://styleguide.yahoo.com/resources/optimize-search-engines/seo-basics

Zhu, X. & Gauch, S. (2000). Incorporating quality metrics in centralized/distributed information retrieval on the World Wide Web. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR '00*, Athens, Greece.