TECHNICAL REPORT

Search Basics - GBY

http://www.cput.ac.za/academic/faculties/informaticsdesign/research/warc

Prof. Melius Weideman 25 October 2016

Note: This Technical Report will be updated regularly. Kindly refer to the date above to ensure you have the most recent version.

1. INTRODUCTION

GBY refers to the worlds' three biggest search engines – Google, Bing and Yahoo! (see Table 1 – this is true if one ignores the Eastern part of the globe). The purpose of this report is to enable the reader to effectively use search engines to find relevant information quickly.

2. SEARCH ENGINE REACH

The Western world use three search engines much more than any others: Google (www.google.com), Bing (www.bing.com) and Yahoo! (www.yahoo.com). Most of the early search engines (Excite, Altavista, World Wide Web Worm and others) have disappeared, leaving Yahoo! to take the lead in the late nineties. Soon after Google was born (around the year 2000), it's logarithmic growth in popularity bumped Yahoo! back into the second place, with MSN (Microsoft Network) in third. Afterwards MSN was renamed to Bing, and currently the popularity of these three leaders are as noted below.

2.1 Western rankings

Search engines are often ranked by measuring the number of core searches done on them per time period. A core search is one search done by a human, executed by typing a search query into a search box on a search engine homepage, and pressing Enter. There are many other types of triggers starting an Internet search, but core searches are easy to measure, hence its use. According to Table 1 (data for USA only), Google was leading in June 2015 with 64%, followed by Bing with 20.3% and Yahoo! with 12.7% of market share. Another way to summarise these figures is to state that Google has almost double the market share of all the others combined.

U.S. Explicit Core Search

Google Sites led the U.S. explicit core search market in June with 64 percent market share, followed by Microsoft Sites with 20.3 percent and Yahoo Sites with 12.7 percent. Ask Network accounted for 1.7 percent of explicit core searches, followed by AOL, Inc. with 1.2 percent.

comScore Explicit Core Search Share Report* (Desktop Only) June 2015 vs. May 2015 Total U.S. – Desktop Home & Work Locations Source: comScore qSearch			
Core Search Entity	Explicit Core Search Share (%)		
	May-15	Jun-15	Point Change
Total Explicit Core Search	*100.0%*	*100.0%*	*N/A*
Google Sites	64.1%	64.0%	-0.1
Microsoft Sites	20.3%	20.3%	0.0
Yahoo Sites	12.7%	12.7%	0.0
Ask Network	1.8%	1.7%	-0.1
AOL, Inc.	1.2%	1.2%	0.0

**"Explicit Core Search" excludes contextually driven searches that do not reflect specific user intent to interact with the search results.*

Table 1: Search leaders for June 2015 (http://www.comscore.com/Insights/Market-Rankings/comScore-Releases-June-2015-US-Desktop-Search-Engine-Rankings).

2.2 Global rankings

However, the picture looks a bit different globally. Searchengineland provides some statistics on the status when comparing the figures from all over the world – see Figure 1 for the top six. Not surprisingly, Google is still the overall leader by far, but further down the list there are some interesting changes. Two big search engines from the Eastern side of the globe feature strongly on the list: China's Baidu (www.baidu.com) and Russia's Yandex (www.yandex.xom) have large user bases, and contribute a fair part to the overall user figures. Also, note that the picture does not change dramatically over time. In a window period of six months, only the last two on the list of six have swopped their rankings once.

.

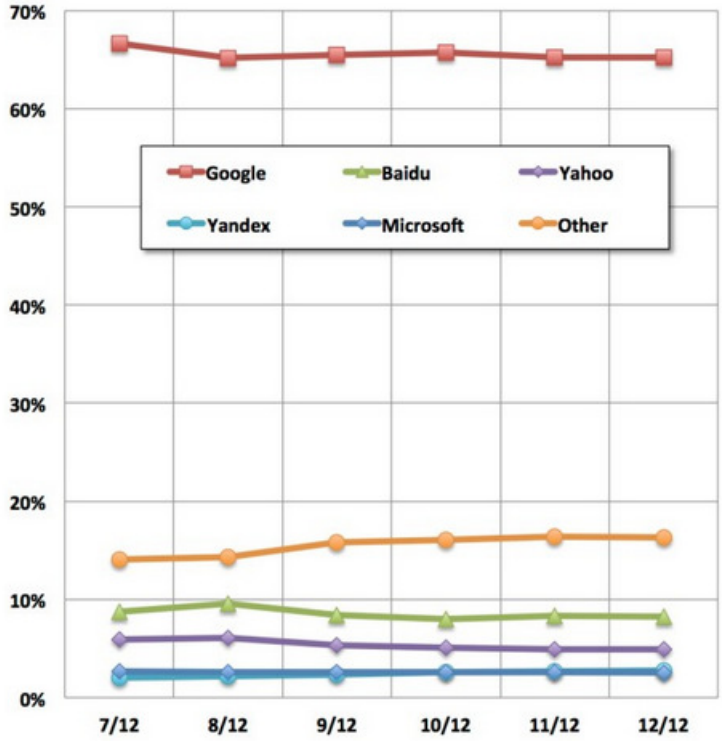

Figure 1: Search leaders – the top six global picture (http://searchengineland.com/google-worlds-most-popular-search-engine-148089).

Since Google and the “Other” category produced rather large figures, the graph becomes difficult to interpret towards the bottom of the scale. When these two higher figures are removed and the graph is rescaled, a clearer picture of the fight happening down at the bottom becomes evident – see Figure 2.

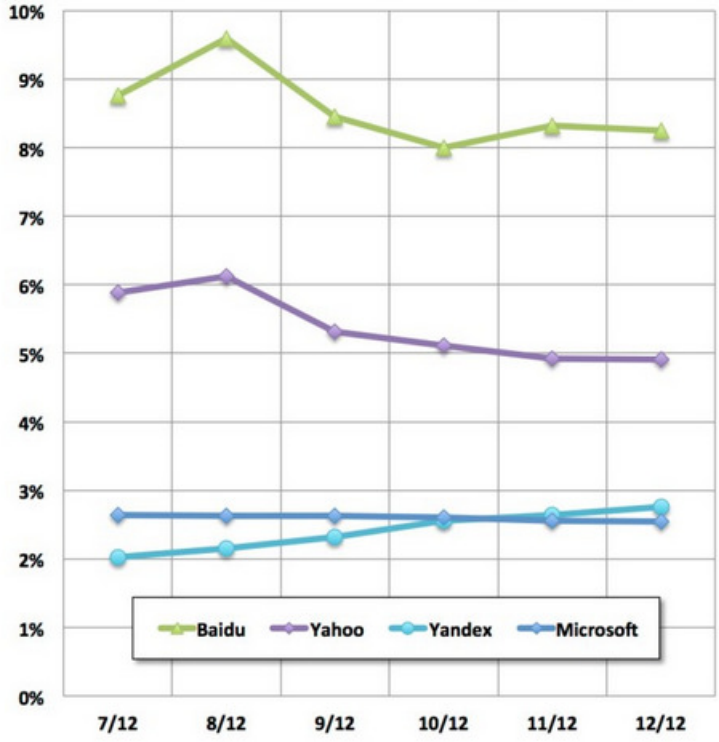

Figure 2: Search leaders – positions numbers two to five (http://searchengineland.com/google-worlds-most-popular-search-engine-148089).

Other useful comparisons are available – as one example, view the 59 slides at http://www.slideshare.net/GlobalAnalysts/search-engine-market-share-qtr-1-2015

To summarize, Google is by far the most popular search engine world-wide, but one should be careful to fall into the trap of using only Google all the time. The other search engines on the list of the top six did not get there by offering low quality results or a small index. They all produce relevant results most of the time.

You are urged to get into the habit of using other search engines from the top six list in parallel with Google – sometimes others provide better results. Most of them provide English interfaces and results.

3. QUERY GENERATION

After choosing one or more search engines to use, one should consider the process of building an effective search query – one which has a good chance of producing relevant answers quickly. Note that most search engines are not case sensitive – typing in: Cape Town or cape town will produce the same results. If you do want to retain the capital letter in your query, you could use the “ operator – see Section 3.4.

Also note that most search engines put an implied „AND“ in between words you type into a search box. For example, if you type in:

latest news missing boeing 2015

the search engine will interpret it as:

Which webpages can I find which contain the words ***latest*** AND ***news*** AND ***missing*** AND ***boeing*** AND ***2015***, all on the same webpage?

3.1 Generating effective queries

It might be easier to first define a „bad“ search query – one which is likely to produce many but irrelevant results. Single-word queries are unlikely to produce good search results, especially if the words are general in nature. Examples of queries which will produce many but probably mostly irrelevant results are:

weather
politics
sport
religion

The problem is that these queries do not really state what it is you are looking for. One should expand these queries by adding more words, thereby focussing more closely on what it is you really need. At the same time, one should use fewer stop words – those with a very general meaning (like: the, that, information, computer, I, need, country, etc). Instead, focus on using more keywords (specific words with one meaning only) as part of a meaningful phrase. After improving the sample queries above, they could be used as:

weather 7 days london uk
president usa 2005
winner 2015 tour de france
difference beliefs muslim christian hindu

So, three things one must continually attempt to do when building search queries, are:

a. Use more rather than fewer words, to focus your query better.
b. Use keywords rather than stop words.
c. Connect keywords in a meaningful phrase.

Remember that the search engine does not know what it is you want to find. It does use many other factors to try and read your mind, in an attempt to give you better answers. These other factors include:

d. Your location (if your mobile device has a GPS, and it is enabled).
e. Your search history (topics you have searched for before might help them guess what it is you are looking for now).
f. Similar topics other people have been searching for lately – Google implements this as predictive search (also called Autocomplete) – see Figure 3 for an example.

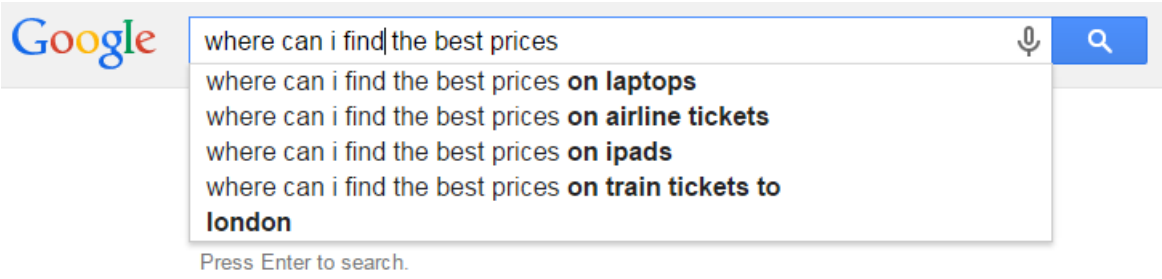

Figure 3: Predictive search results (www.google.com)

Predictive search attempts to guess ahead what you might type next, by getting ideas from common search queries of other users. Not only does it save you typing (one can select an option without typing the last few words), but it also provides you with other ideas of how to construct your query – sometimes better than what you could have provided yourself.

Another final hint to assist you in effective query building: where possible, include any one or more of the following to focus a search:

- figures,
- meaningful abbreviations,
- non-English terms,
- technical phrases,
- brand names.

Below are some examples of basic queries: first an inefficient one (which will probably not produce relevant answers), followed by the same query which has been adapted using some of these ideas.

martin luther king sound
1963 speech martin luther king i have a dream audio

bicycle chain shifter
mtb shimano chain derailleur

learn guitar chords
a b c d e f g chord sheet guitar tutorial

I want more information on eating disorders
"eating disorder" anorexia bulimia bed coe symptoms

engineering studies at a university
civil mechanical electrical engineering degree europe university

3.2 Local search

One of the strongest current trends in search is so-called „local search". It is based on the fact that millions of users globally rely on their mobile devices (since it is convenient and always on) to supply answers to their information needs. If you are on the sidewalk of a city, and want to buy a pizza, you might use your mobile device to search for the closest pizza restaurant. However, if you are in New York, the location and telephone number of a pizza restaurant in Paris will not be helpful. Hence the tendency of search engines to incorporate your physical location in the algorithm to determine the best results for you. Often the results are supplied in order of distance from where you are, making the final choice easier for the user.

The Uber taxi service also relies heavily on the searcher's locality (and that of the taxi drivers) for providing the best search results. Some digital cameras also store the co-ordinates of the location where the picture was taken (see Figure 4), to enable many other useful features.

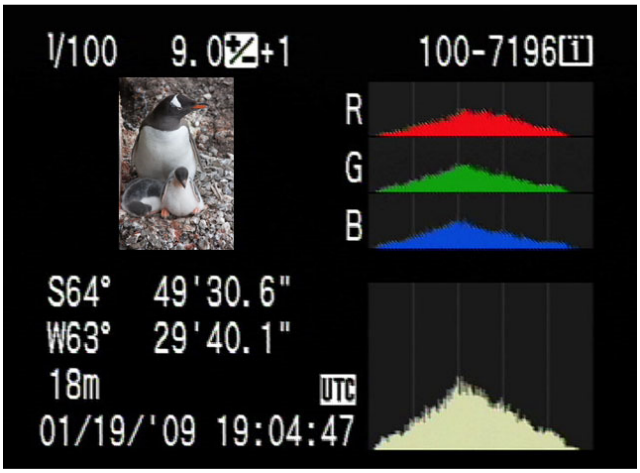

Figure 4: GPS co-ordinates stored with a digital photograph (www.sjphoto.com/gps-canon.html)

3.3 Voice search

Another fairly recent development is the use of voice search instead of typed queries. It has become popular since it removes the typing interface, which is cumbersome and time-consuming. However, it does introduce a new variable into an already complex situation, in an attempt to make life even easier for the information searcher. This new variable is the effectiveness of the software on the mobile device, which has to convert the speech to text. From there onwards, the searching takes place as normal, with the search engine producing results on the screen.

The concepts noted under Section 3.1 above hold equally true for voice searching. Only now, one has to ensure to pronounce the words carefully, just as you used to ensure that you spell the words correctly while typing. Furthermore, one's accent might also affect the quality of the results. Most systems were created by English-speakers, so any user with a "non-English" accent might have to take extra care when pronouncing search queries using voice search. Some systems use training to get the conversion software used to a specific voice, thereby increasing its accuracy.

3.4 Search operators

Most search engines offer a number of ways you can further focus your search, thereby improving the quality of your results. Some examples are given below, in no particular order, for using with Google. You should try your own versions of queries based on these operators, and inspect the answers on the SERPs (Search Engine Result Page) to confirm that you can see that they have affected the search results. Most search engines provide help menus and/or "Advanced Search" links to assist you with these operators.

a. Specifying words to be in the URL. ***allinurl:***

Starting your query with this operator will give you results of webpages which contain your query words in its URL (not necessarily in the body text).

Example: ***allinurl: faq nikon camera***

Try this example, and then look at the URLs of the answers on the SERP – they should all contain the three search words in the query, as part of the URL.

b. Specifying results to contain a certain price only. ***$:***

Attaching the dollar sign to a value will limit search results to prices of that value, associated with the other words in your query. It will include prices specified in other fiscal units – not just US dollars (no space after the colon).

Example: ***price new car $:30000***

Try this example – all the websites on the SERP should contain the exact price you have specified.

c. Looking for definitions of concepts only. ***define:***

When you precede your query with this operator, you will get mostly definitions of what you specify (no space after the colon).

Example: ***define:error 404***

Try this example – all the websites on the SERP should contain a definition of the 404 Error.

d. Exclude certain words from being part of the answers. ***-***

Sometimes your query has multiple meanings, and you want to suppress or exclude the meanings you do not want in your answers. You could use the "exclude" operator: dash (**-**) (no space after the dash).

Example: ***manchester -united city map***

Try this example – none of the websites on the SERP should contain the phrase "Manchester United". Note that this operator is the "short dash", not the "long dash". The long dash will not perform this function.

e. Only a certain file type should be in the answers. ***filetype:***

This operator enables you to specify that you only want a certain file type to be part of the results on the SERP.

Example: ***latest fashion jeans levi filetype:jpg***

Try this example – all of the websites on the SERP should contain at least one image (with a .jpg extension), which is related to the keywords in your search query.

f. A phrase instead of a keyword must be in answers. ***"***

This operator enables you to specify that two or more words, in the exact sequence, must be on the page of the answers you want

Example: ***lyrics jackson "I'm bad, I'm bad"***

Try this example – all of the websites on the SERP should contain the phrase "I'm bad, I'm bad", in this exact sequence. In other words, websites with the words "I'm" and "bad" - not next to each other - will be excluded.

g. You want to restrict the search to only a given website ***site:***

This operator enables you to specify that the terms of your search query must only be searched for in the domain that you specify.

Example: ***job web designer site:www.careerjunction.co.za***

Try this example – all of the websites on the SERP should have a variation of the URL you have listed, and will be relevant to the search terms you have noted.

h. You want to broaden your search using a wildcard. *******

This operator enables you to specify that you want to include words derived from those in your query.

Example: ***winter olympics 20****

Try this example – all of the websites on the SERP should contain the words “winter olympics” (not in sequence though), and a year number starting with “20”. In this way you exclude those with the year starting with “19”, but include all events, current and future, between the year 2000 and 2099.

i. You want to force your browser to use a non-default country ***/ncr***

This operator enables you to specify that you want the search engine index of a specific country, not the country where you are at the moment (non-country-redirect).

Example: ***www.google.de/ncr***

If you type the above without the /ncr, your browser will probably ignore the .de (Germany) suffix, and load www.google.com if you are in the USA, www.google.co.za in South Africa, etc. Using this operator will force the loading of the German Google instead of the default.

3.5 Query fine-tuning

One should get used to the idea that the first search done for a topic is not always successfull. The best way forward is to simply continue the search process, after altering one or both of the following: the search query, and/or the search engine. The search engine is easy to change - just use any of the other big search engines‘ URLs specified earlier, with the same query (initially).

Improving the search query could be done as a process of successive refinement. After every unsuccesful search, increase the query length by adding one or more keywords. The number of results on the SERPs is likely to decrease, which indicates that you are focussing your search – this is good news!

For example, you might start a search for a topic with the query:

iphone 6 price

You will probably be unhappy with the large number of results. So try this next:

best iphone 6 price cape town

Again you might recieve too many answers (but fewer than before), so refine your search even further:

best new iphone 6 price free delivery cape town

This process could continue until you are happy with your results.

In general:

- **The LONGER your search query, the FEWER results will appear on the SERP.**
- **The FEWER results on the SERP, the BETTER – it means you have focussed your search well.**

4. SERPs

After typing in (or speaking) a „good“ search query, you will now be faced with a set of results – often termed a SERP. Some examples are given in Figure 5.

Figure 5: Sample SERPs from big search engines (www.bing.com www.yahoo.com www.baidu.com)

Search results are listed on SERPs with the most relevant result first, and the least relevant result last. However, the search engine does not know exactly what the user wants to find, so it has to guess at the sequence when compiling this ranked list. The closer the search engine is to ranking the results to your opinion of how it should be done, the higher the degree of relevance.

It has been proven in research that most users prefer reading only the first few results, or at most the first page (typically 10 results). This tendency has major implications for e-commerce and other similar websites, where the ranking of a company, product or service on SERPs determines their financial success.

This drive for ranking highly on a SERP has spawned a set of techniques (and a big industry using them) called SEO (Search Engine Optimisation). When SEO has been implemented correctly, it will increase the ranking of (a) given webpage(s) for one or more key phrases, by listing those webpages high up in the natural (organic) search results (the main result block on the SERP). SEO involves fine-tuning some of the elements on a webpage, and some outside the webpage, to ensure favour from the search engines for a given topic (once that page is indexed).

PPC is one scheme that involves payment to the search engines for placing advertisements. This time the ranking of the advertisements is based on the amount offered by the advertiser for a given keyword (or key phrase), and other factors. An example of PPC advertisements is given in Figure 6. Not all search engines offer a paid placement scheme like PPC. Advertisements are normally listed at the top of the screen, and, if more than a given number appear, the rest will be down the right hand side of the screen. PPC requires that the advertiser writes a short advertisement, chooses a bid price for one or more keywords/key phrases, sets up an account, and then releases the advertisement to the search engine. When any user now types in a keyword/key phrase that closely matches those of the advertisement, the ad might appear on the user's screen. If the user clicks on an ad, he/she will be taken to the specified landing page. As with natural results, the higher up the ad appears, the more likely it is that it will attract clicks, and the more the advertiser stands to make. The higher the bid price the advertiser is prepared to pay, and the higher the quality of the landing page, the higher the ranking of the advertisement will be.

Either SEO or PPC (Pay Per Click) can be used to improve rankings of websites on search engines.

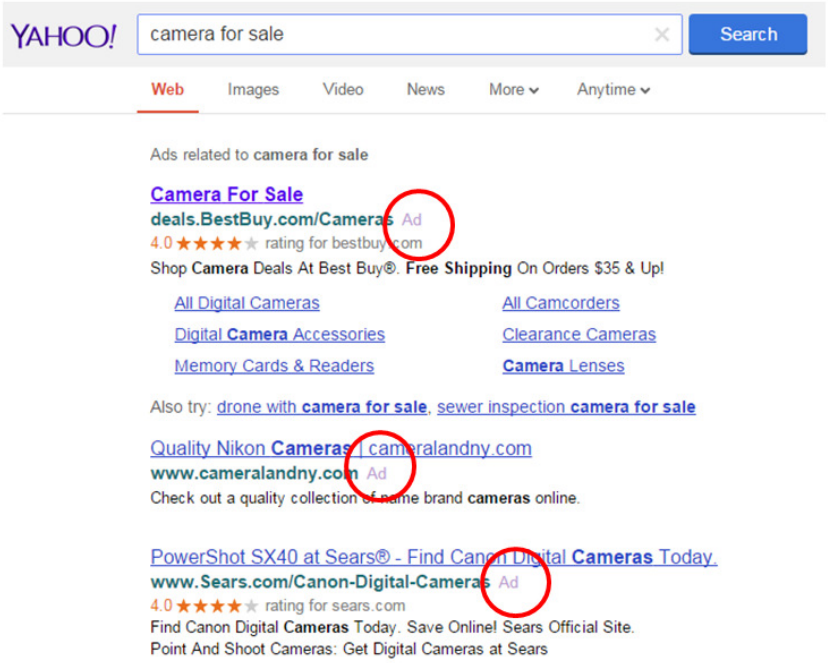

Figure 6: Example of PPC advertisements (www.yahoo.co.za)

There is a close link between the user generated query, the keywords of the PPC advertisement, and the keywords used on a webpage. **In summary: the more accurately you specify your query, the better the chance that your results (both organic and paid) are close to what you are looking for.**

5. MSS (MULTIPLE SIMULTANEOUS SEARCHING)

As noted before, searching for information is not always successful on the first attempt. Apart from changing the query and or the search engine, one can try another approach: simply drill down a bit deeper on the first one or two SERPs, instead of giving up after checking only the first one or two results. Using the built-in multi-tasking functionality of your operating system will help a lot in this process.

For example, you might type in a search query:

second hand mercedes b-series cape town

Since you have started with an efficient query, you scan the first SERP, and believe that the answer you are looking for is actually somewhere amongst the first 10 results. Your SERP might look like the one of Figure 7.

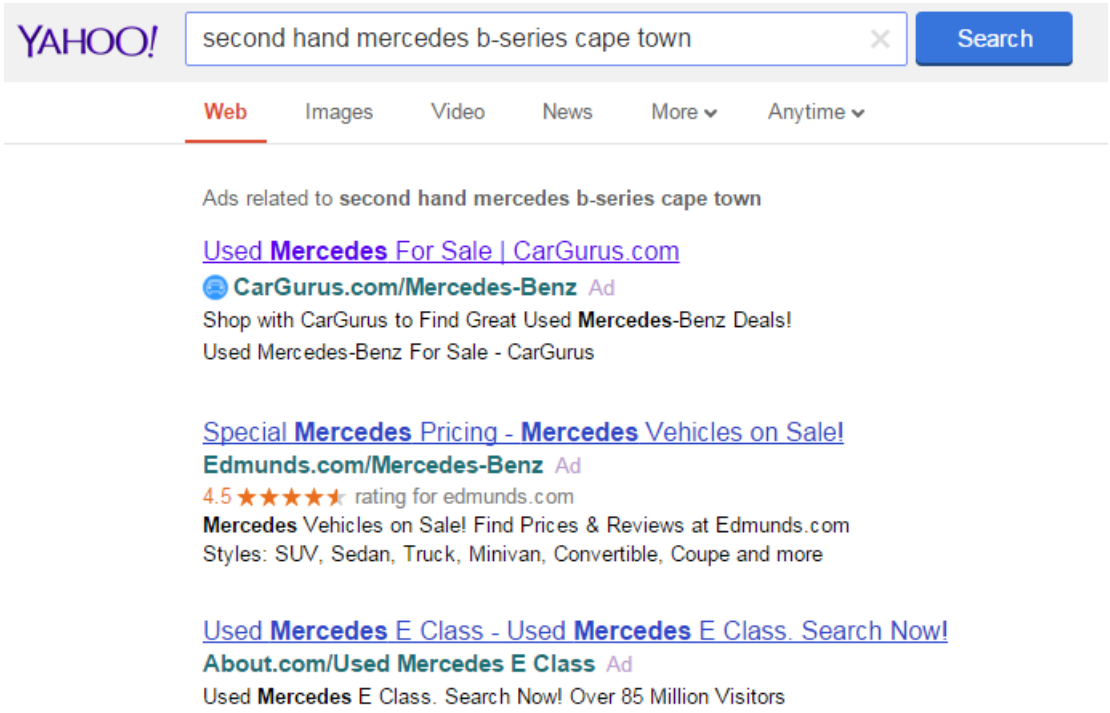

Figure 7: Example search query and SERP (www.yahoo.com)

Some websites are slow to load, so if you approach this process in a serial way, you might waste time – loading, waiting for and inspecting one website after the other. You might as well load all 10 the answers listed on the SERP in quick succession, thereby allowing the slow websites to load while you are spending the waiting time loading the others.

Proceed as follows (this is much quicker to do than to read…):

a. Right click on the first answer, select to open it in a new tab.
b. Right click on the second answer, select to open it in a new tab.
c. Right click on the third answer, select to open it in a new tab, etc.

By the time you have opened all 10 websites (a few seconds later), the first few you have loaded will be ready to view. Your browser tabs (at the top) could now look like Figure 8.

Figure 8: MSS: Multiple open windows (www.bing.com)

Now it is an easy task to simply click on the first tab and scan the website contents. If you think it is irrelevant, close the tab. If it looks relevant, leave it, click on the second tab and repeat the process. Within a very short time, you will end up with a few open tabs, all of them webpages that are likely to provide you with the required information.

The author believes that if one works through all the examples in this document, the reader will find that he/she is more efficient at using search engines, and understands the process of creating "good" search queries better.