

INFORMATION TRACKING IN THE INFORMATION AGE (SLIGHTLY) REVISITED (2003)

INTERNET AS A SOURCE FOR MARKETING INTELLIGENCE

The basis for this article is another article which I published online in 1999 (www.sam-mag.com) . A lot has happened on the Web since that time; search engines have merged, others have almost disappeared, and new research tools have emerged such as Google for instance. At one point in time, I even believed that it was no longer necessary to write anything else on this subject because Google had become so perfect that so many of us could find anything they wanted on the Internet. All you had to do was to type the words you wanted and Hey Presto! Here came the information. However, in the past few years, new events have occurred. Multimedia search engines and capabilities have been dramatically improved, therefore changing the way one may track information on the Web. A flurry of information websites has cropped up all over the world. The Google search bots are now overwhelmed with requests from web sites and webmasters who want to register with the famous search engine. As a result, Google nowadays is no

longer able to keep track of all the information that is available online and the once very useful meta search engines which had fallen into disuse are now back on the agenda. Besides, referencing websites is no longer totally free. Or rather, it still is but it is also possible to pay for better referencing. If an organisation pays the price it will be able to place an ad or a link to its URL on the topmost lines of any search results page, even if the information to which it relates has little to do with what the user has typed in the search box of the search engine. This can be really annoying and what's more, these so-called sponsor links can hide the actual results of your search. Thirdly, it seems that what we used to call intelligent agents have definitely thrown in the towel, as I had suspected 5 years ago. So now it's back to basics and to the good old search engines that we all knew. And that takes us back to this article written in 1999 whereas the Internet was only budding. Already at that time I was pointing out that the importance of what you were

searching (the information itself as opposed to the envelope) was at the top of the agenda. And this has not changed. When you come to think about it, all the new gimmicks that have been invented have changed absolutely nothing to the fact that if you do not know what you're looking for, then you will never be able to find it. This is what triggered my decision to re-view this article and to produce a slightly revisited version of it. I really believe that the simple advice in this article, based on common sense will still be helpful to the reader who wants to improve his or her ability to find information on the Internet.

In the first version of this article I mentioned one of the early versions of Copernic. New versions have been issued since then. Myriad functionality has been added to this tool, but I am not sure it changes anything either. Of course you will get nicer buttons and sleeker menus and the usability will have been improved (actually even that is not certain). But the overall way it works has not changed very much.

Last but not least, I have reformatted this article into HTML in order to improve its visibility and readability.

INFORMATION TRACKING IN THE INFORMATION AGE RE-VISITED

The so-called *information age* may well not be the age of information after all but that of *data* instead, for it is not easy to find the valuable information you are after on the Web. Meta-search engines,

intelligent agents, search-bots, web spiders and other Internet buzz-words are as many obstacles to the understanding of information-tracking on the Internet by ordinary users. However, you do not need to be an IT-expert, far from that, to find interesting things on the Net, whether they be of the professional or recreational kind. Indeed, nowadays, it is possible to include the Internet as part of one's strategy for Marketing intelligence; even if it is still early days to replace all your information sources by the Internet, one can surely rate the Internet as the best place to *start* when you want to launch a Marketing Intelligence campaign.

THE DISCRIMINATING INTERNET USER

Yet it is becoming increasingly difficult to find the right things on the Web. When I started publishing my material online at <http://visionarymarketing.com> in the beginning of 1996, there were 3 Marketing-related sites in Yahoo! Nowadays, there a hundreds of them scattered amongst several dif-

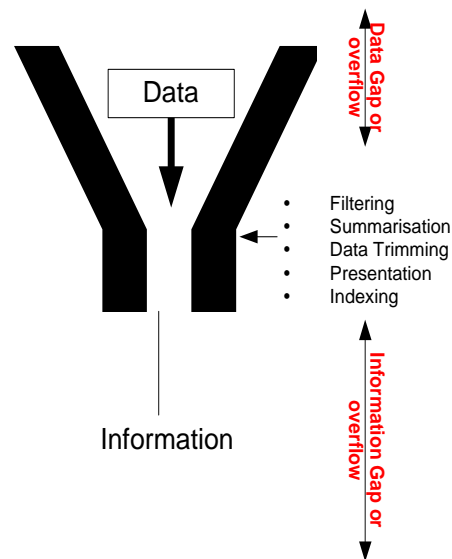


Figure 1 : The Information funnel (1)

ferent categories. As a consequence, it is more and more difficult to have a new site referenced within search engines and there arises two problems from that. On the one hand, for already well referenced subjects it implies that we have to face a surfeit of information and on the other hand, there is a lack of information concerning certain other domains. So that the information seeker is always confronted with either too much data to sort out or too little to present. Later in this article, we will learn that the methodology for finding what you want has to be dependent on how popular the subject is. Without developing complicated theories about the nature of information, we need to understand what information really is before one searches for it.

Will the excess of information available on the Net satisfy our average user? Not really, because most of this information is *irrelevant* to him or her. In fact, it cannot really be depicted as information as it is a mere aggregation of data. In order to make matters clear, one can represent that accumulation of data on the Internet with a funnel (see Figure 1). At the top end of the funnel is raw data, which comes unrefined from the Internet. At the other end, comes the information, i.e. the data that has been evaluated, checked and sorted out; in order to become useful information, the data needs to be checked for relevance, usefulness, and for its latest update date(s). Despite the endless Marketing gloss about so-called *intelligent agents*, you will find that there is still a lot of work left to the people for discerning data from information and that is good news for the future of the human race. However, search-engines are incredible tools and provided you have a decent Internet connection and a reliable ISP (*Internet Service Provider*), you will find them amazingly quick at browsing through the entire mass of websites that they index (I have gathered some valuable tips about search-

engines in Box A).

It would be too easy if everything that has been posted on the Internet were true. Yet how many times have we heard it said in the media that it has to be true since it was published on the Internet. Of all times, people have used the means of communication in their possession in order to inform, but also convince, seduce, misinform or even deform the truth. This is true of speech, newspapers, TV and all other means of communication. The Net is no exception to this rule. It serves nothing to deny this most obvious trait of human nature. One has to fight it when it becomes dangerous and threatening to our liberties and the rest of the time we have to be careful and decode the information properly when we receive it. The only defence we have is our discrimination of what is true, false or suspicious and to refuse to believe data blindly when it has not been checked. Obviously, it is not always possible to ensure that the information is true, but then it's recommended to ensure that its sources are quoted to allow cross-checking examinations and also to ensure that some serious research has been carried out. At the end of the day, very little else than your own discriminating eye will be telling you what's wrong about a piece of information. So, you may also want to show it to experts of the relevant subject and take their advice for it. Only the discriminating user will be able to make the most of this invaluable information source which is the Internet.

SAVVY SURFERS

A savvy surfer, and what's more a professional who spends his life dealing with information will surely have his own special ways to find the right information very quickly. However, good information researchers do not forcibly follow any precise guidelines. Some of them are better than others and mainly so because they can afford to be sufficiently *patient* and re-

arrange or refine their keywords long enough in order to get to the right Websites. They surf the Web patiently and probably less methodologically than would seem necessary and yet they do so with an open mind and they know that data that is only vaguely relevant to the original subject may well prove to be the indication that *information* is there at the end of that dark tunnel. This is what makes them different from ordinary surfers for they are above all patient curious and open-minded.

Deriving a proper methodology from that is therefore rather awkward. On the other hand, other researchers – maybe a little less patient – can also do a pretty good job by dint of personal organisation and because they can choose the right *keywords*. In fact, I have observed that information-tracking is best when both of these profiles – the patient and the impatient types - are brought together in a team.

Somehow, one is under the impression that all this is not for *Joe public*. Thus, if you do not have that much time on your hands you definitely need a *check-list* or at least a method which will help you out of the Internet maze. My aim here is no other than try to *help* Internet users by providing precise, down-to-earth advice and also showing examples which can be used as training sessions.

'NODAL POINTS'

In his 1996 best-seller *Idoru*, William Gibson's hero Laney, a professional information researcher was looking for 'nodal points' in order to gather the data he needed. He could barely explain what it was. All he knew was that he sensed that among the vast amounts of raw, shapeless data that he was studying there were moments where what he called 'nodal points' emerged, that is to say where the data seemed to gather some meaning and started to form infor-

mation in a stricter sense. All he had to do was to go from one *nodal point* to another in order to find the information he needed. This led him to the discovery of things that others would not even dream of finding.

To those who think that this is pure science fiction and bares no resemblance to reality whatsoever, I would like to tell that in fact, it is a lot like the best way of finding information on the Internet today. William Gibson - although he started the 'cyber' craze when he invented the word 'cyberspace' in *Neuromancer* - confessed recently that he had never used a computer let alone the Web when he wrote his best-selling title. However, I would add that he understood more about information-tracking that have millions of Internet fans before him who are struggling more and more to find what they want on the network of networks.

Let's find out how we can use Gibson's allegory to become those

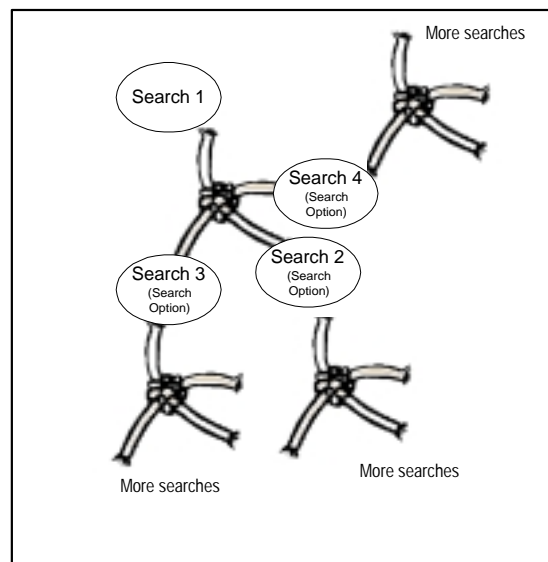


Figure 2 : Nodal points

outstanding information trackers who will impress our bosses and/or clients.

Nodal points are the points where your request comes to a halt and branches into other directions. Each time you encounter a nodal point, break down your search into as many topics as possible and start searching again by redefining or rearranging your keywords. Each topic (let's call them *search options*) has to be taken individually for better results. The nodal points methodology will be explained by ways of an example.

SAMPLE SEARCHES

First of all, let me explain how to organise your search. All along, I will resort to examples to drive these points home. Please refer to the search-tools box whenever necessary.

We have established earlier that there are two main kinds of subjects :

- Those where information is rare (either with remarkable or no remarkable keywords),
- Those where information is plentiful, i.e. related to very popular subjects where the level of *noise* (that is unwanted data mingling with the ones that are relevant to you) is rather high.

The result is that either you do not get the information you need, or it takes forever to find it because you do not know where to start. Even if it seems to be a bit of a catch-22 situation, you should never let it take you down and in any case you should never bother about the number of web sites involved with a search. It might be daunting to know (Excite will tell you for instance) that 150,000 URLs match your criteria, but in fact you don't need to browse all 150,000 pages to find what you need. Above all, you need to stick to what you *find* and forget all about what you could have found but haven't. This would be counter-productive, and let's face it, who would be able to read 150,000 pages anyway, be it on the web.

UNUSUAL SUBJECTS

I have deliberately chosen a weird subject in order to show how powerful the Internet has become. 3 years ago, searches on the same subject produced practically no result. Still, information about *Simeon Stylites*¹ is rare. Check your favourite reference books, and you won't find much about it. Actually, even online, this is where we are going to start. There are quite a few good reference sites online (do refer to the BOX A : INFORMATION TRACKING ESSENTIAL TOOLS box for details).

This will be step 1 of our process, i.e. checking the spelling and writing down the keywords. A request through <http://infoplease.com> (select the *all sources* option) will almost immediately bring a summary bio of Simeon Stylites : “ [Gr., = of a pillar] died 459 ?, Syrian hermit. He lived for more than 35 years on a small platform on top of a high pillar. He had many imitators



Figure 3: St Simeon Stylites (left), shown at the moment he was called down from his pillar by Church authorities.

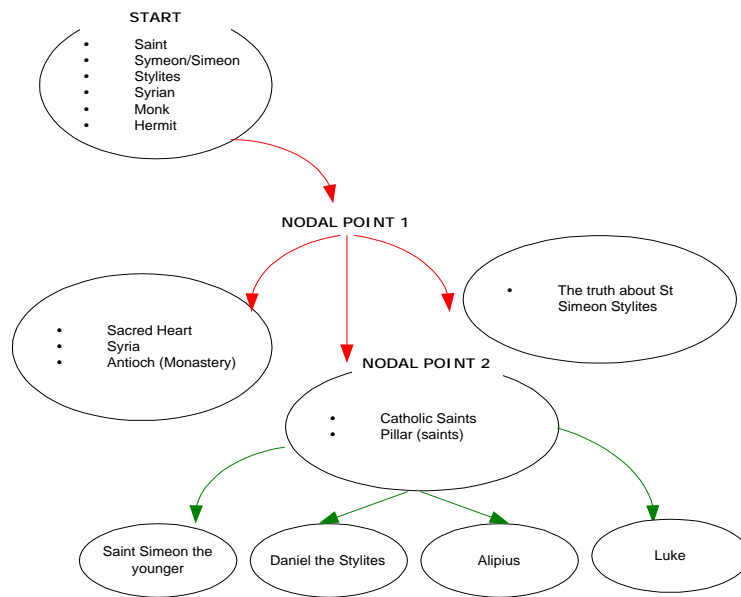


Figure 4 : Nodal Points by example

(called stylites) and gained the reverence of the whole Christian world. Feast: Jan 5.” is the entry at Simeon Stylites of the Columbia encyclopaedia 5th edition dated 1993. Although 1993 is already a long time ago now, there are few chances given the subject that new facts about Simeon had been gathered in the meantime.

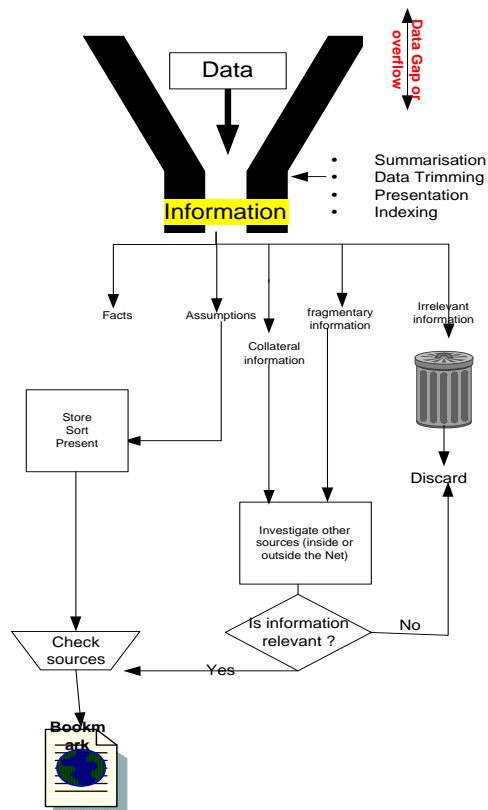


Figure 5 : The Information funnel (2)

APPLYING THE NODAL POINTS METHODOLOGY

We will therefore take this information as it comes from *Infoplease.com* and we will deduce from this cyber-cyclopedia a few keywords which will lead us to our first ‘nodal point’. You will notice that there are two possible spellings for Simeon (Symeon being the alternative) which means that - if we were to come across some difficulties for finding any information at all with the *Simeon* keyword, we could use *Symeon* instead. That leaves us with 5 or 6 keywords which are relevant to our research :

1. Saint
2. Simeon / Symeon
3. Stylites

4. Syria / Syrian
5. Monk
6. Hermit

We won't need more to start with, so it is needless to seek more keywords at this stage. In a sense, this is already too much for if we were to launch a keyword search with so many keywords, then it would be likely that ordinary search engines would return very few results.

As a consequence, we will restrict our search to the three keywords which seem most relevant to our request, and we will enter them directly into the input box of one of our favourite *meta-search engines*¹, i.e. metacrawler. We will stick to the first three words which by the way form a valid phrase (*Saint Simeon Stylites* is indeed better than 'Stylites saint Simeon' for instance or any other combination of these 3 words). Metacrawler will then return a few pages which in their turn will give us a few hunches about the Saint and his life. As a consequence, we will learn from our first wild shot that Simeon was part of the *ascetical* saints and that he had a few pals named *Daniel*, *Simeon the younger*, *Alipius*, *Luke*, ... and that he had strong connections with the *Antioch Monastery*.

What it teaches us is that we are very likely to find more information about St Simeon if we were to search for *Luke or Daniel the Stylites* instead of Simeon only. Pointing at the *Antioch Monastery* might prove useful in the long run too. This is the sign that I have reached a nodal point in my search. From then on, I am going to launch more requests in each of these individual directions (see Figure 4).

Once I have explored the relevant nodal points¹, I can then order my results in a bookmarks file in which all the websites I have found will be presented. I will have ordered the links and grouped them by theme; I will also add a brief description under or beside each link to increase the legibility of the bookmarks page. although there is – theoretically – no end to a search, one could consider it done when the *need for information* is actually fulfilled. One could complete the version 1 of our *information funnel* (Figure 1) and then describe the process which has taken us from *raw data* to *information* per se (see Figure 5).

POPULAR SUBJECTS

Popular subjects are numerous on the Net. *Free downloads* are a staple for instance, but then I guess that you have found a way of finding them already. Thus I will concentrate on a subject which is a little more relevant to our readers : *Change Management*.

This subject requires a very different search strategy altogether in that sense that neither of the two elements of this combined word stand out sufficiently. Hence I will proceed as follows :

a)Change or Management cannot be treated as two separate keywords, so I will enclose them into a pair of double quotes which indicates that *Change Management* is an expression (i.e. *one* keyword only and not a combination of keywords),

b)Search for *links* on "Change management" ("management of

change” is a likely alternative) through the main meta-search engines. A page of links is a good place to start with. Why waste time searching while others have done it for you already,

c)Track the “*Change Management*” Yahoo! category. If you don’t forget to include the double-quotes, Yahoo! should return (amongst other things) http://dir.yahoo.com/Business_and_Economy/Companies/Books/Shopping_and_Services/Booksellers/Business/Titles/General_Business/ with a hyper-link to reach that category. Reading this page very quickly, you will find <http://www.utsi.com/wbp/reengineering/> which offers a list of Internet resources on the same subject (<http://www.utsi.com/wbp/reengineering/resources.html>). Then you can move from link page to link page until you find the information you are after. This will save you the trouble of building these pages of links yourself, and then you will have more time to *qualify* these links and find what is relevant in them.

d)Use Copernic to browse the various search engines, throughout the web, or by using one of their predefined collections of search-engines in Copernic 98 plus (News, Business, French search-engines ...)

e)Refine your search by making it more specific, i.e. by pointing your engine at *Change Management toolkit* or *documentation* or *check-list* or *definition* etc . Your results sheets might be more specific; if they are too specific, change the keywords until you find one that brings more matches.

SAMPLE SEARCHES

We have listed a number of

techniques which can be of help when launching a Marketing Intelligence campaign on the Internet. Obviously, and however helpful these few tricks may be, one will still have to develop the basic skills of a good Information seeker which are the ability to read fast and mainly to tell the difference between useful data and that which has to be discarded. Even so, searching is an iterative process.

I wish that these few tips will convince you that you should not be put off when you launch a quick keyword search and you get no results, and I hope that thanks to this article you will understand that it may be worthwhile to invest (and not waste) a few minutes or even hours in a quest, which eventually will lead to a wealth of information when it would have taken a few days to find the equivalent – or worse – in your public library. And all this at your fingertips.

I think it’s worth giving a thought or two.

Note ¹ : You do not need to search all the nodal points that you have found. Once one or two of them have led you to the desired result you must consider your job is over. Only the objective of your search matters; this method is just meant to be a support tool and is in no way an end in itself.

Note ² : see BOX A : INFORMATION TRACKING ESSENTIAL TOOLS for details about meta-search engines and other search tools.

Note ¹ This choice is purely arbitrary and was not made to promote any particular point about religion, let alone support Simeon’s Stylite’s views, which with hindsight appear a lot less saintly than their utterer.

BOX A : INFORMATION TRACKING ESSENTIAL TOOLS

At least one thing is certain, it is that a daunting list of search engines (<http://searchenginewatch.com/links/article.php/2156221>) will not help you much. Unless you feel particularly at ease with detailed search techniques, I advise that you stick to a few online tools which I will detail hereafter. The following list is therefore not comprehensive. Most of these tools are referenced at <http://visionarymarketing.com/pointers.html>.

ONLINE DIRECTORIES

e.g. yahoo! , Msn (2003 note: Netscape Netcentre is now powered by Google and is no longer an online directory)

They work wonders if you are in a hurry and if you know in anticipation what category you are after. But don't abuse them. Let us assume that you are looking for all the available websites dealing with Internet Marketing for instance, then <http://www.yahoo.com> (or .de for Germany / .co.uk for the UK / .fr for France ...), are the right starting points for you. Also check Excite at http://companyinfo.excite.com/site_map/index.html or Cyberexpress at <http://www.cyberexpress.com>. A flurry of specialised online directories are also available such as the business directory at <http://www.the-business-directory.net>

INDIVIDUAL SEARCH ENGINES

e.g. hotbot, lycos, altavista etc.

There are so many of them that it would be impossible to list them all here. What's more, you won't need them much once you have discovered their betters, i.e. the meta search engines. However, some of them are better than others, simply because they are more popular. Search engines require that webmasters (that is, those in charge of the Internet site) put a request forward to them in order to be referenced. Then they launch what they name their *spiders* on these URLs in order to index whatever they can from the home page or the collection of pages that has been registered. What it means is if a webmaster is too lazy for registering his URLs with all the main existing search engines, then he or she will stick to the most famous of these engines. And this is why information tracking is more accurate on the big than the small ones. As a result, in most cases, you will not have to bother about smaller search engines, and mainly if you are a beginner, you run better chances by trying to optimise your searches on one of the three main engines. Please note too that Lycos is mostly helpful because it allows you to search for multimedia files, altavista is good for searching newsgroups and particularly helpful for its babelfish (<http://babelfish.altavista.com/>) online translation service. I don't think much of its translation capabilities but you will find it invaluable if you want to browse in a language which you master only imperfectly.

ONLINE REFERENCE BOOKS

e.g. infoplease, onlelook, Merriam Webster's, britannica online, Msn Expedia and Encarta etc.,

Reference books such as the Encyclopaedia Britannica online are not free. But may try it for a period of 30 days. I recommend Infoplease (<http://www.infoplease.com>) which comes free of charge and never fails to deliver the basic information that you need. Encarta works good too (<http://www.encyclopedia.com>) Onelook (<http://www.onelook.com>) will give you free access to all sorts of online dictionaries including technical ones. So, if you after a medical or engineering term, you will easily find them there. Merriam Webster's online (<http://www.m-w.com>) is *the* reference for US-English, it's free and it comes with a free online thesaurus and complete etymology. It is an absolute must on the Net.

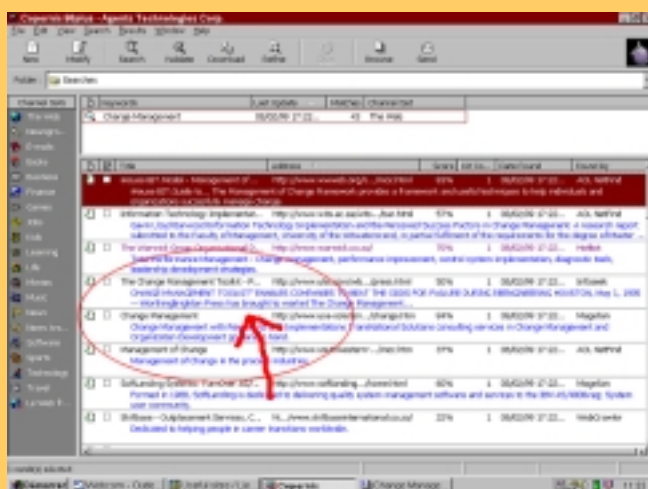
META SEARCH ENGINES

Very little is known of meta-search engines and yet I barely use anything else. On <http://visionarymarketing.com/pointers.html> you will find a list of some of the Net's most interesting meta-search engines. Three meta-engines are head and shoulders above the rest :

- (1) Metacrawler (<http://www.metacrawler.com>) is very straightforward and it now includes a multimedia search engine (2) Mamma (<http://www.mamma.com>) is as straightforward as metacrawler (3) Kartoo (<http://www.kartoo.com>) literally maps your search and displays the links between the websites which are tied to one another through their keywords. Although developed by a French team, versions in several foreign languages including English are available. There are also options which enable you to restrict your search to a specific country.
- Please note that Inferencefind, my one-time favourite has now sadly gone under.

MULTIMEDIA AND FILE SEARCHES

Searching for multimedia files is no longer a challenge. Almost all search engines including Google cater for multimedia files. Searching for different file formats is also possible now. When in Google, select the "Advanced Search" link at and in the file format box select any of the available file formats available such as Word, Powerpoint and Excle,...



OFFLINE SEARCH TOOLS

Nowadays, the *intelligent agent* craze seems is definitely over and it's back to basics (hence this article).

Nonetheless, one of the best search-tools available is a Canadian offline meta-search engine called *Copernic*. It is available from <http://www.copernic.com>. It offers close integration with Microsoft Internet Explorer and lets you refine, modify and sort your results sheet. It is one of

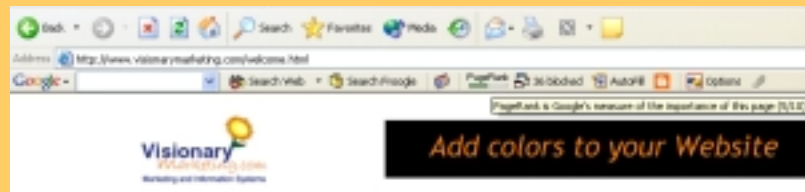
the essential tools for Marketing Intelligence on the Net.

ONLINE BOOKSTORES

<http://amazon.com> or Barnes and Nobles (<http://www.barnesandnoble.com>) are good places to start if you wish to build a bibliography. Direct links to Amazon's description help you add meat to a website or report effortlessly, and if you are a site owner and wish to make money with it, you may join the Amazon partnership program.

RANKINGS

Google now provides excellent rankings thanks to its Google toolbar. I find it most useful when I want to see whether a Website is really popular.



RECOMMENDATIONS

- The last thing you need to do is to try all the search engines one by one. You will find that serious Internet information trackers have their pet search-bots and that they stick to them. Some will try to impress you with a list 2 feet long full of exotic Urls but you do not need to believe them. It serves no purpose at all,
- Do bookmark the *advanced* page of your favourite search or meta-search engine, even if you don't use their options all the time. These advanced pages let you enter restrictions for your choices, or they offer a selection of file types (multimedia for instance),
- Unless you are an Internet search wizkid, don't bother about boolean search options (AND, +, OR, ... or any combination of those) as most of the time they prove useless and cumbersome gimmicks,
- Switch engines if you find that the connection is too slow (or if you have given up hope on your favourites) but otherwise avoid changing tools all the time. Performance varies greatly from day to day as always on the Net.
- Most users stop after the first page of results. That's a big mistake because search engines accuracy ratings are, most of the time, wide of the mark. This is mainly because *spiders* use some special hidden Html tags (called meta-tags) which most webmasters fail to update or even create. As a consequence, the presence of their website at the top of the search result list is rather uncertain, even if their relevance is maximum. So please keep on reading, even if result lists most of the time look like garbled lines of endless rubbish,

Beware of online directories, for they resort to individuals who index the sites which they arbitrarily sort within categories. As a result, some of these *portals* prove very difficult to use and they will not show *all* that is available on the Net regarding your search. If the indexing staff at Yahoo! for instance have chosen not to display a website it won't show at all even if it contains exactly the information you require.

BOX B : BASIC PRINCIPLES FOR EFFECTIVE INFORMATION RE-SEARCH

ORGANISE YOUR TOOLBOX

Organising your toolbox is a must if you want to search effectively. You might use the *bookmark* capability of your browser but you might well find it unwieldy and difficult to port from one browser to the other. Moreover, if you travel a lot like me and need a constant access to a bookmark and/or change pc's very often you will soon find that a local bookmark or favourites file is not enough. Either upload your bookmark file to a free online server space like *geocities* for instance or log on to <http://visionarymarketing.com/pointers.html> to use mine.

CHOOSE YOUR KEYWORDS

This is where a true information research campaign begins. Do follow these simple guidelines for better results :

Your keywords have to be unambiguous. The word *Utilities* for instance, is ambiguous because it refers either to a Public Utility Company or a Computer routine. The Internet being a major purveyor of such computer programs, one had better avoid the term *utilities* when one searches for information about electricity. It is then recommended to replace that keyword with "public utility" (double-quotes mean that the 2 words form an expression) or with other more specific keywords such as power or electricity, gas, "environment services", etc. depending on what you are after. In other words, choosing the keywords also forces you to refine the *objective* of your quest,

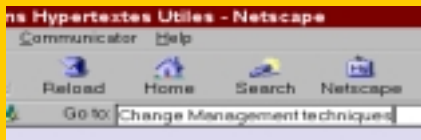


Figure 6 : Smart Browsing

Use a combination of 2 or 3 keywords (but not much more than that). Entering just one may prove ineffective while retaining 4 of them will more likely display an empty results page,

Include the words *Link* or *Links* in your search in order to track those pages where others have done the hard work for you already,

Use double quotes to force the engine to find an expression : "*Teddy bears*" will show information on Teddy bears, while *Teddy Bears* will lead you to *Teddy or bears* (or both in some cases),

Most search engines are supposedly case sensitive and it is said that they will be able to tell the difference between *ABC* and *abc* for instance. In practice, this is not always true. Check the case of your keywords if you experience difficulties, otherwise pay no attention to it.