

Information Overload

Volume 5: Issue 51, November 2006

Welcome:

As an information professional I have been asked many times to justify the existence of the Information Centre or Library where I have worked. Management wanted to know why they should continue to pay for literature searching, articles and Inter Library Loans when they could find everything they needed "on the web". They also wanted to know why they should continue paying me when everyone had access to the Internet through their desktops.

As most of you will be aware, the inroads being made by Google and others into making published works available on line, and with virtually every library catalogue available for searching, the arguments that we have used in the past may appear to be wearing a little thin. So how can we – as Information Professionals continue to justify our existence, and our salaries in the face of mounting digitisation and availability of free (and not so free) information on the web?

We would like to thank you in advance for forwarding this edition onto friends, colleagues and other interested readers. Please note that all back issues of this edition, as well as our registrant resources edition can be read and/or downloaded from our web site – <http://www.iea.com.au> should any of the topics be of interest and use.

Lorraine Bradshaw
Marketing Coordinator and Projects Officer

In this Issue we will be looking at:

- How big is "the net" anyway?
- Just the facts ma'am!
- Where did you say you found it again?
- Fee vs. Free
- A Thought to ponder.

How big is "the net" anyway?

It appears there are no definitive figures for the size of the Internet. Which is hardly surprising given the speed in which web sites can be updated and deleted. And then of course there are those web pages that are hidden behind firewalls, inside databases and frames based web sites. It is interesting that these kinds of website cannot be accessed by the many robots and spiders that crawl through the world wide web...if they come to a link that is broken, or the site requires you to ask a question in order to retrieve an answer (think Whitepages and Amazon) they simply move on. Which is probably a good thing in

© IEA 2006. All rights reserved. You are free to use material from the Information Overload newsletter in whole or in part, as long as you include acknowledgement of source.

Information Enterprises Australia Pty Ltd
Unit 4, Upper Level, 201 High Street, FREMANTLE WA 6160
Tel: 08 9335 2533 Fax: 08 9335 2544 e-mail: training@iea.com.au

some cases. Can you imagine the fun and games of finding telephone numbers and names and addresses in your search results?

But going back to the size of the “net” for just a minute. The figures that are quoted cannot say with certainty whether these figures are based on total web sites or the number of pages within the websites. Perhaps a slightly better indication of the size of the WWW is the number of active Domains there are in operation. Typically a domain name is registered for a short period of time (typically 1-2 years) after which time it can be renewed, should the owner want to. And if they don't want to renew then of course the domain is up for grabs to the quickest new purchaser.

However, even this is not a true indication of active websites as some domain names are purchased and never used.

Is it any wonder then, that information you found a few days ago may not be available today. Or if it is there, how certain can you be that it is the same information. I.e., it hasn't been changed by the owner or editor of the site. The answer is – well unless you are the owner of a particular site, you usually can't tell from a cursory glance.

The problems caused by this rapid evolution and the world of the information provider are immense. But it certainly beats the old cliché, that everything is on the net...oh and if it is on the net, then it must be free to use and utilise. How can I put this politely – wrong and wrong again.

So how can we as information specialists persuade the powers that be (and pays our wages) that not everything is actually on the Internet, and if it is, then:

- a) it must be factually correct.
- b) it can be found.
- c) and it doesn't cost anything to access.

Just the facts ma'am

Taking the first point we mentioned. Not everything that is provided on the internet can be classed as factually correct. Any website can be biased. It all depends on where you are looking from. Why should I make such a rash statement. Well – everyone has their own point of view. Everyone has their own set of reference points and value systems in place. What may seem like a perfectly logical set of events may in fact be just someone's recollections. Whilst it can be argued that sites such as Wikipedia (the online encyclopedia) is one of the best “free” websites for just about anything, readers should also be aware that the information contained in the website is open to edit from just about anyone. It is also interesting to note that in the FAQ's and editors pages, that some of the “editors” were objecting to other “editors” work and demanding the persons work be removed and the person prevented from “defacing” the site again. Interesting.

But it emphasises a major point – most websites do not have peer reviewed material, let alone edited material contained in the pages. Load up your favourite search engine of choice and try spelling a word incorrectly and see how many references you get. Oh and for the seriously bored amongst you – try <http://www.alltooflat.com/geeky/elgoog/m/index.cgi> - yes the search engine backwards and see what you get when you type in a word or two.....some people really do have far too much time on their hands.

Apologies for the digression. But it does make our lives rather interesting don't you think? (BTW – my children told me about that one!!)

© IEA 2006. All rights reserved. You are free to use material from the Information Overload newsletter in whole or in part, as long as you include acknowledgement of source.

Information Enterprises Australia Pty Ltd
Unit 4, Upper Level, 201 High Street, FREMANTLE WA 6160
Tel: 08 9335 2533 Fax: 08 9335 2544 e-mail: training@iea.com.au

But going back to the serious side of what makes a “fact” a “fact” and how can you tell? Well what would happen if the web master of an organisation had a rigid set of beliefs, would that person be willing to add information to the site it considered to be against his/her own beliefs? I suppose it depends on the person and the individual organisations, but it does go to show, anyone with an agenda can change the face of a website, but because it is “on the net” people are more likely to believe the information contained in its pages.

A hint – don’t. If you are basing business decisions , or even life saving ones, on only one or two pieces of information I would suggest that your analysis will be a little “off”.

Where did you say you found it again?

As you know, information hidden behind firewalls, passwords and contained within databases form what is known as the Deep Web. As opposed to the listings we receive from the many search engines we utilise, which has been termed the surface web. Imagine an iceberg, and the bulk of the material is hidden below the surface, well that is exactly the problem we face when we are trying to find information on the net.

The surface of the ocean which forms the barrier between what you can see and what you can’t is like the many web sites that require you to input a password or do something before you are allowed access to the inner pages of the web site. Firewalls, password protected sites, any site that requires you to ask a question (amazon is a very good example), or old web sites that use frames based technology – our own current web site <http://www.iea.com.au> is a good example of this kind of site (thankfully a new one will be coming soon).

However, there is yet another problem associated with this kind of site. Not only do the robots and spiders fail to index the sites (all they see is a dead end), so they do not rank high in the search engine results,

Even if the website does not require passwords etc, your website may not rank highly on the search engine listings because your web site does not contain sufficient content for the creatures to stick around. If that content does not contain any of the search terms being looked for by the many surfers out there, the spiders will pass you by. Your site should be content rich and full of relevant keywords. However, this should not be the excuse you need to go and add a million variations all on the same theme as this is the realm of the spammers. If the robots and spiders think you are trying it on, you will be relegated to the depths of the search engine listings, destined to forever eat, canned meat products. Give information of relevance and use and people will find you...eventually. It has to be said that a Google ranking for most sites is about a 3 or 4. Anything better and you have had to have been around a while, or you have paid a considerable amount of money to get to where you are. What that also means is – you have to train people to go beyond the first one or two pages of results to find information of potential interest.

By the way – the first listings are those paid ones we mentioned. The ones that follow tend to be .org; .asn; .edu and .gov sites rather than the individual company sites, or the sites that end with tags such as .info. But then as information professionals you probably already knew that – sorry.

And of course the final problem we will discuss today, if the information is on the web, you found the site, and you found the information you were after, the other major stumbling block to getting information of interest, relevance and use is cost. Most organisations are in

the business to make money after all, and are not going to release premium content for free if they can charge users a fee for accessing the material. Granted that cost may only be a small, but imagine if everyone within an organisation pays the small figure listed on the web site for the item they are interested in. If there is no central acquisitions department/person, or the information is not collected and catalogued before being passed on to the person who requested the information then it is likely that duplication of material and therefore cost is likely to occur.

Why do organisations need a librarian anyway? We can and do save organisations money!!

Fee vs. Free

Imagine then the dilemma facing many information professionals. Teach people how to surf the net properly and outsource yourself out of a job? (Actually unlikely given the rate of turnover of most organisations), or emphasise the need to have a central point for lodging information, so that other people may utilise the information individuals have downloaded and therefore save money.

In order to be effective and accountable, business decisions need to be made using the best information that is available, and that means advising our clients and colleagues of the best places to find information. And if there is a fee attached to the information then so be it. Because, whilst some people think that the best information is free information, as we mentioned before – why give it away, when you can sell it, and sell it many times over!

But we have gone ahead of ourselves. The first step in the debate over free vs. fee is where is the information stored?

Some of the best information resides in the many libraries around the world. Whilst the move has been towards putting library catalogues online, do you still have to go to individual libraries to check to see if a library has a particular copy of something? – well not anymore. For instance, OCLC has made access to 10,000 library catalogues available for searching through their world catalogue project – <http://www.worldcat.org>

However, there are limitations. Worldcat may give you the libraries that hold a particular title, it still does not mean that you can gain access to a particular title unless you are a member or subscriber to the service. Take for example the book I searched for – “As a man thinketh” by James Allen. This book is out of copyright (the author died in 1912) you would think then that libraries would make this title available to anyone who wanted to read it. Not so.

So – not content with having to wander down to my local library I used the Google books search option - <http://books.google.com> - here you are given two options, search the holdings of a title or search for full text items.... Same book, same problem. Limited access, no full copies of the book available....

Google says “Right now, most of our books are provided either by publishers and authors, through the Google Books Partner Program, or by our library partners, through the Library Project.

The Partner Program is an online-book marketing program designed to help publishers and authors promote their books. Publishers and authors send us their books and we digitally scan them and make them findable in our search results - all for free. For the most part, you can browse through a number of pages of these books, typically about 20%.

The Library Project involves partnerships with several libraries to include their collections in Google Book Search. Books scanned through the Library Project are displayed to you like a card catalog, including basic info about the book, and in some cases basic info plus a few snippets - sentences of your search terms in context. When these books are in the public domain, we'll display the full text of the book, from start to finish."

Note the last point – when these books are in the public domain, we'll display the full text of the book, from start to finish....well I know the book is in the public domain...is it available – no. I am aware however, that digitising and making available every title that is now in the public domain is a major task, and whilst Google is making inroads into the vast quantity of material – it will take some time to get to the one that I am interested in. So now what, I still want to read it, just don't want to have to buy it!

As always there are still a couple of options open to me – a search on the internet brings me a site that is dedicated to this little book - <http://www.asamanthinketh.net/> but do I want to make someone else rich? – the answer is no. Whilst I can download a copy of the book – I have to provide information regarding "me" so they can bombard me with sales pitches for other items they have on offer, no thankyou. So I turned to perhaps the first and best site for public domain works on the World Wide Web. Project Gutenberg, <http://www.gutenberg.org> and keyed in the same search and sure enough – James Allen, As a Man thinketh....

I am aware that this only helps if the item you are interested in, comes in the form of public domain works. What happens if you want articles or papers? Well there are some sites that allow you to access this kind of material – <http://findarticles.com> is one, which may help. It is said to contain 10,000,000 articles from thousands of publishers..... some of which are available free of charge, but a majority of these articles are listed as premium content only...

And what of Google scholar? – <http://scholar.google.com>

Well a quick search of Google Scholar may give you the references, but not much else. Even those items that say they are downloadable – doesn't mean the items are free or available to download through Google at all.... Another major issue with Google scholar is that it only indexes things that it has "permission" to index. Any individual, institution or organisation that does not want it's item(s) listed anymore can have the item removed.

Frustrated yet?

Some of the best information is available from the world wide web, but as we have found out over the years, so is some of the worst, and we have to trawl through hundreds of references to get to the ones we think may be of interest and use to us. We could go directly to the better sources, but that presupposes we know where those sources are located on the net, and then we still may not get access to the material because we don't subscribe to a particular resource. It also presupposes that we know how to search the many sites that contain the information that we are interested in. Whilst some sites do use some form of Boolean logic to assist us, a lot of sites still do not have the capabilities provided by the search giants of Dialog, STN and Lexis Nexis.

If you have limited time, and want quality information without having to visit the individual publishers websites, then these are an excellent option. But they're not cheap. But if you want the information quickly and don't have time to spend trawling the net – it could be argued these kinds of sites are priceless.

So does free really mean free?

Of course not.

- You have individuals spending time looking for information....the old adage "time is money" is still valid. In most cases these people have no idea where the best sources of information are. They start and stop their search with Google, and yes we are all guilty when it comes to that score.
- Information is being duplicated, costing time and money; and
- Is poor decision making – based on incomplete information gathering costing your organisation time, money and more?

As you can imagine this is just a brief foray into the realms of the Internet debate, but as always we hope we have given you some food for thought.

Kind Regards

Lorraine

A Thought to Ponder:

"If there was nothing wrong in the world, there wouldn't be anything for us to do"

George Bernard Shaw, 1856-1950

Irish dramatist, writer and critic

Your comments and suggestions on the subject of this newsletter are most welcome. Or if you would like to see other issues covered in future editions, please email me at training@iea.com.au. Please feel free to pass on this newsletter to your colleagues' friends and associates. To subscribe they should send an e-mail to training@iea.com.au with "subscribe newsletter" in the subject line.

If you have any suggestions as to what should be included in future editions, then please send an email to training@iea.com.au. If you would prefer not to receive this newsletter, please send an email to training@iea.com.au with "unsubscribe newsletter" in the subject line.