

The Internet as a Source of Relevant High-Quality Information

Karen Blakeman

RBA Information Services, Reading, UK
karen.blakeman@rba.co.uk

INSOURCE 2008: Conference on Professional Information Resources for Business,
Management, Marketing and Research
Prague, February 5-6, 2008

Abstract

Anyone and everyone can publish on the Internet if they so wish. Not all, though, are qualified or knowledgeable in their chosen subject and some even publish deliberately misleading information. Web 2.0 technologies such as blogs and wikis can add to the confusion. Separating fact from fiction, and scrutinising the credentials of an author are now an essential part of any research. One does not have to pay for reliable information and there are many good quality free resources. Being able to quickly recognise these is a key skill that all researchers must acquire. This presentation will look at the steps one can take to assess the quality of resources and their trustworthiness, and the weak and strong features of both priced and free information on the Internet.

Introduction

Information of questionable quality did not start with the Internet or Web 2.0. Errors and deliberately misleading information have always been with us, and we have never needed electronic networks to help us invent or perpetuate inaccuracies and urban myths.

We all know that when Sir Isaac Newton was hit on the head by an apple he came up with the laws of gravity. That story was first published in an essay by Voltaire, long after Newton's death and there are several other versions of the story, but it is unlikely that the apple actually fell on his head.^{1 2}

Then there are the experiments from the late 1800's saying that spinach is high in iron. This myth originated in 1870 with Dr E von Wolf,³ or was it in 1890 with Gustav von Bunge? The figures remained unchallenged until 1937 (or so it is said), when it was discovered that the content was 1/10th of the claim.

Moving into the 20th century, in 1912 it was claimed that an unknown form of early human was discovered in a gravel pit in Sussex in the UK. 'Piltdown Man' was later discovered to be a hoax.⁴

Some of us who were growing up the late 1960's can remember the press coverage of Paul McCartney's marriage to Linda Eastman who, we were told, was a member of the Eastman family, which in turn was part of the Eastman-Kodak photographic empire. That story was started by a British journalist from an unverified assumption. Other journalists and newspapers simply copied the 'fact' without checking it.⁵

More recently in 2003, the New York Times reporter Jayson Blair resigned his post after it was revealed that he plagiarised and even invented interviews and stories.⁶

Looking for quality on the Internet

As has already been said earlier, poor quality information has always but the new technologies do make it easier to initiate, maintain and propagate it. The temptation is to use the first 'googled' article or document that has a recent date stamp on it and which appears to come from an authoritative source. But date stamps can be misleading, typographical errors occur all too easily, and even the Financial Times has in the past confused stock market charts for the BOS (Bank of Scotland) with those for the RBOS (Royal Bank of Scotland). Web site updates can result in documents that are several years old being labelled with a recent date, and people make mistakes. It does not matter whether the source is free or priced: quality assessment should be applied to both.

So how does one go about assessing the quality of a site or page? There are a number of criteria you should look at but the importance and weighting that is attached to each of them varies depending on the type of enquiry and on how you plan to use the information:

- is the information time-dependent - are you looking for the most recent news on a topic or is it a precise, factual query, for example the turnover of a company
- are you looking for authoritative data, peer reviewed sources?
- are you looking for differing views and opinions on a topic, product or company?
- is an expert assessment of an industry required?
- is it 'nice to know' background information or mission critical, such as supporting a recommendation to go ahead with a corporate acquisition? How paranoid do you need to be?

Typically, one would like at:

- how you found the site – was it 'googled', found on an evaluated listing or recommended by a trusted friend and colleague?
- the publisher or provider of the information
- the author, editor or compiler of the information (may be the same as the publisher)
- date of publication or update
- if it is an information aggregator, what sources are covered? Are there in any gaps in the time series?
- geographic coverage
- industry sector coverage
- time span – historical information may be important to you
- can you verify the information via an independent source or trace it back to the original source
- if uncertain about the veracity of a site, try seeing who links to it

Many of the above are self explanatory and with priced services the information is either included in the document itself or can be found in the database description. Where the fee based services are sometimes lacking is when it comes to source lists for news, market research and industry reports. Because of the length of the lists, they are time consuming to update and so may not accurately reflect current coverage. Gaps in the chronological coverage are not always recorded either. Free services, on the other hand, often do not have a source list at all, for example Google News.

Two criteria that often cause difficulties for researchers when using free resources are identifying the publisher or provider of the information, and the date on which the information was compiled or updated. If one cannot easily identify the provider, it could be argued that it should not be used at all. However, that web page may be the only document containing the vital data you need.

The URL and domain name of a web site can be easily checked. The structure of the web address can give clues as to the type of organisation. Some types of domains are restricted, for example those that end in .gov and .gov.uk can only be used by US and UK governments respectively and .ac.uk is for the UK academic sector. For a list of country and type of organisation domains see the MozillaWiki TLD List at http://wiki.mozilla.org/TLD_List.

The ownership of domain names can be found by looking in the domain name registers. Allwhois (<http://allwhois.com/> MACROBUTTON HtmlResAnchor <http://allwhois.com/>) and Domain Tools (<http://whois.domaintools.com/>) are just two of many starting points. It is possible for a person or an organisation to hide their identity by using an intermediary. If that is the case and there is no means of identifying the real owner on the web site, the information the site contains should be regarded as 'suspect'.

The date of publication or update is usually clearly stated in books, journal articles and newspapers. Priced collections are generally reliable and easy to search and sort by date, but even highly regarded sources such as The Financial Times may not be so easy to use if you opt for the free version.

When it comes to a web page, even if there is a date it cannot be always be trusted. It may be the date when the information was loaded onto the server, not when the information was gathered or written. Worse still, some pages are constructed in such a way that they always display today's date. If there is no date on the page at all, when the page is displayed type the following into the address bar of the browser:

```
javascript:alert(document.lastModified)
```

Whether or not this works depends on the web server and how the pages have been set up. It may always give today's date, date "unknown" or the default year of 1970. Even if it does give a sensible date, do remember that it is the date when the page was loaded or reloaded onto the server and not when the content was written.

Firefox users can use the Tools, Page info option to view date written or updated, but it is subject to the same drawbacks as the javascript:alert command

In contrast, articles on wikis are automatically stamped with the date and time of the last update. Look at any Wikipedia article and click on the history tab. That will tell you when the page was last modified and give you a history of all the previous modifications; you can even view earlier copies of the page to see how it has changed. Similarly, blog postings are date and time stamped.

Web 2.0

Web 2.0 is a concept. It is a way of thinking and working, the most important aspects being collaboration and sharing. Web 2.0 allows any number of people to author, edit and update information, and to share views and opinions on a subject. The technologies can also

generate 'mashups' where information from two or more disparate resources are combined to generate a new one.

The concept of sharing is not new: discussion groups, web forums and email lists have existed for a long time. The collaborative resources that we now have, though, far exceed them in numbers and usage. This is advantageous if you are looking for a wider range of viewpoints and blogs in particular are a valuable source of opinion on companies, markets and industry sectors.

Blogs

Yes, the vast majority of blogs are irrelevant as far as serious business information is concerned, but authors, researchers, analysts and publishers are using these to provide news and comment on their sectors. Blogs may be single or multi-authored and, if enabled, the comments option encourages discussion. The Oil Drum (<http://www.theoil drum.com/>) is an excellent example of an industry sector blog with serious debates taking place in the comments area.

Before using a blog, check the following:

- who is/are the author(s)?
- is there an author profile or 'about us'?
- the bias of the authors
- date of last posting and frequency of posting

Wikis

The most famous wiki is Wikipedia (<http://www.wikipedia.org/>), which has attracted controversy over its 'all can edit' policy. In fact, that is not totally true: some pages are locked so that only certain people can edit pages and this is a feature common to many wikis. For example, Alacrawiki (<http://www.alcrawiki.com/>), published by content aggregator Alacra, has a collection of industry overviews that anyone can view but only Alacra editors can change.

When using a wiki as a source of information check:

- who is the main publisher of the wiki?
- is the purpose and coverage of the wiki clearly stated?
- is there a way of identifying who has made changes to a page?
- are some or all of the pages locked or restricted for editing?
- use the history tab to see when a page was last updated, what has been changed and when

Mashups

Mashups combine information from different sources to create a new presentation of that information. The process is not new. For example Yahoo Finance gathers information from stock market, news and company information databases to generate a single profile on your specified company. The quality of mashups is dependent on the quality of the source databases and on how the information is combined. If the source is inaccurate and out of date then the mashup will be as well.

Even if the incoming data is high quality, the mashup process may create errors or omissions. I can search Google Maps for local businesses and as well as a list of the companies I can view their location on a map. This is a typical mashup and in the UK the source for the business listing is Yell.com. If I search Yell.com for restaurants in my area I am given a full list. If I repeat the search in Google Maps, Google invariably omits 2 or 3 restaurants from the list – and the omissions change from day to day! Something in the ‘mashing’ process is losing information but not in a predictable or consistent way. This type of error may not be critical depending on your application, but it is one to bear in mind when using mashups.

Finding quality information

Looking for shared experiences of previously discovered sources is often the best way to locate quality information. Email discussion lists have long been used for this. For business information BUSLIB-L (<http://list1.ucc.nau.edu/archives/buslib-l.html>) and the AIIP mailing list – Association of Independent Information Professionals - (available to AIIP members only) are two of the best. Today, there are social bookmarking sites such as FURL (<http://www.furl.net>) and Del.icio.us (<http://del.icio.us/>) where researchers tag, annotate and share information on documents that they have found useful. Also, look for blogs that review and compare services.

If you are new to a sector or need to broaden the scope of your research, look for evaluated listings and subject guides. These rely on human input rather than computer generated mashups. One such example is Alacrawiki Spotlights (<http://www.alacrawiki.com/>), which is divided into industry sectors. Each page identifies reliable sites that provide statistics, news, directories and market information in that industry. Alacra also uses another relatively new technology in its Alacrasearch (<http://www.alacra.com/alacrasearch/>). This is a Google custom search engine that runs your search on business sites selected by Alacra. Neither Alacrawiki nor Alacrasearch are comprehensive – they are not intended to be – but they do help focus your search.

Stock markets and share prices

Current share prices are easily found free of charge. Yahoo Finance, for example, covers what it calls the ‘major stock markets’ of the world, but there are many sites, including the stock markets themselves, that provide individual country coverage. Free share prices are delayed by 15-20 minutes and historical data is often only available as charts. Yahoo does offer some free historical information for download but it varies from market to market. Charting and analysis options are also usually limited on the free sites.

Company financials

For companies listed on the stock market, annual reports and accounts are often to be found free of charge on the company’s own web site. For other types of company, the official company registers are the best starting point. See <http://www.rba.co.uk/sources/registers.htm> for a list of some of the registries by country. The data may be completely free of charge, or only basic information such as name and address is supplied with detailed data available on subscription. Some registries provide the information in the local language only and archival material may be limited or non-existent.

Free services are sufficient and usable if you are looking at just one or a handful of companies. If you need to quickly identify companies operating in a particular sector and

across several countries then the priced services win every time. Multi-criteria searches for companies by sector, product, country, turnover, no of employees etc are just not possible using free sources. The priced databases also offer more download formats and analysis options. Nevertheless, it is advisable to check the 'cut-off point' of the database: does it have every company that has officially registered in a country or does it only include those above a certain size, for example those with a minimum turnover of €20,000,000? A second point to remember, is that multi-country services may standardise or harmonise financial data so that when you are searching you are comparing like with like. Once you have identified your target companies, you need to be able to access to the original figures.

News sources

Although most newspapers, magazines, journals and trade press have a presence on the Web not all of them provide free access to all of their articles. Some sites may have only a selection of today's articles, and others may charge for access to archives. When they started out on the Web several years ago many national papers offered their content free of charge. Then some started to charge for archived stories and so-called "premium" content. That trend seems to be reversing with many reverting to 'open access'.

Although a vast amount of news is free, tracking down **relevant** articles can be extremely difficult, especially if you are going back several years. The structure of the subscription databases is simply not there. Some publishers have even removed the ability to limit your search and sort by date!

If you are looking for a specific newspaper ABYZ News Links (<http://www.abyznewslinks.com>) and Kidon Media-Link <http://www.kidon.com/media-link/> are two good starting points.

If you want to search across many sources, Google News (<http://news.google.com/>) is a very popular tool for current events. It has good country, regional and industry coverage and the country versions give prominence to local publications, for example <http://news.google.cz/>. There is no source list, though, and coverage is limited to the last 30 days. Some country versions have links to Google's Archivesearch (<http://news.google.com/archivesearch>) on the results list. The implication is that you can extend your current free search to include the same publications going back for more than a hundred years. You cannot! The sources in Google Archivesearch are very different, many of the mainstream newspapers are not there, and a large proportion of the articles are priced. Another approach being adopted by some users is to set up a Google custom search engine for selected news sources. For those of us in the UK, Chipwrapper (<http://www.chipwrapper.co.uk/>) carries out a search on 15 free UK news sites. As it does not use Google News you pick up older material but there is no date search or sort.⁷

Free news search works well for recent stories with very specific or unusual key words, and if comprehensiveness is not required. If you need archival material, want to focus on a company, limit by sector or country, and date limiting options are essential then the priced services are the best.

Statistics and market data

Free official statistics abound and if, for example, all you need are total production figures for an industry there are trade and manufacture associations with overview market data. For statistics two starting points that list sources are OFFSTATS (<http://www.library.auckland.ac.nz/subjects/stats/offstats/>) and NationMaster - World Statistics, Country Comparisons (<http://www.nationmaster.com/>). OFFSTATS includes

sources offering free and easily accessible social, economic and general data from official or similar "quotable" sources, especially those that provide both current data and time series. In the country lists, these are mainly Web pages provided by statistical offices, central banks and government departments and agencies, whereas the topics list is comprised of links to the statistics pages of international organizations and associations and a few commercial sites. Nationmaster repackages information from many different sources and you can compare data in a variety of ways. For example you can select your own countries, a region, or an economic group such as OPEC and then a category and statistic for that category.

When it comes to market and industry research that provides market share information for products and companies, most of us are accept that we will probably have to pay for that level of detail. You may be lucky, though, and find some pieces on free pages. Some government departments produce market reports to support the export activities of companies in their country, or produce detailed analyses of an industry as part of their regulatory activities. To help you track these down there are some interesting new initiatives that are starting to pull these and other free reports together.

ReportLinker (<http://www.reportlinker.com/>) claims to be "a market research engine which provides full access to 1.2 million open access market research reports." Although you can search this database free of charge you have to pay to view the free documents.

Free Research (<http://www.free-research.com/>) allows you to search its database of reports and view the documents free of charge. The drawback is that the search only looks at the title of the report.

ResearchWikis Free Market Research (<http://www.researchwikis.com/>) says it is offers "free Market Research for the Global Community". All the data is publicly available information and the sources of the tables and figures in the articles are given. Editors have to register before they are allowed to create or alter existing pages. The major downside of this site is that many of the reports are heavily US biased.

Conclusion

There is a wealth of good quality information on the Internet, and a significant proportion of it is free. Knowing how to confirm the accuracy, currency and veracity of that information are skills which are easily learnt and applied. Tracking down the individual resources is another matter: it takes time and effort to locate and pull the data together. For simple enquiries such as a company press release, the latest stock price for a company, or today's news on the price of oil, free services work well. When it comes to complex, multi-source, multi-criteria searches, and results that need subsequent detailed analysis, the priced services are easier, quicker and far more cost effective.

Further reading

The Truth, the Whole Truth, and Nothing but the Truth: Error and Misinformation in Media. Paula Berinstein, *Searcher*, September 2006, Volume 14 No. 8, pp 38-52.

The Skeptical Business Searcher: The Information Advisor's Guide to evaluating Web Data, Sites, and Sources. Robert Berkman. CyberAge Books 2004. 312 pp. ISBN 0910965668

1

Falling Apple Story

http://www.sfu.ca/physics/ugrad/courses/teaching_resources/demoindex/mechanics/mech11/apple.html. Page updated 15 December 2003 23:26:52. Accessed 20 December 2007 09:37:41

² Isaac Newton - Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Isaac_Newton. Accessed 20 December 2007 09:47:36

³ Spinach - Wikipedia, the free encyclopedia. Web Page. URL <http://en.wikipedia.org/wiki/Spinach>. Accessed 3 January 2008 16:03:29

⁴ The Piltdown Men, Ronald Millar, paperback: 272 pages, publisher: Flamingo. New edition 21 Mar 1974, ISBN-10: 0586081348, ISBN-13: 978-0586081341

⁵ Internet Beatles Album - Beatles Myths. Web Page. URL <http://www.beatlesagain.com/bmyths.html>. Page updated June 6, 2004. Accessed 3 January 2008 15:04:33.

⁶ New York Times in shock as reporter's lies are uncovered | Media | The Guardian <http://www.guardian.co.uk/media/2003/may/12/pressandpublishing.usnews>. 12 May 2003. Accessed 20 December 2007 09:30:12

⁷ Karen Blakeman's Blog Chipwrapper - Search UK newspapers <http://www.rba.co.uk/wordpress/2007/12/29/chipwrapper-search-uk-newspapers/> 29 December 2007