



## WoW!, Terkko, and IngentaConnect



Although I am still not an unconditional fan of blogs because of the very high noise ratios, there are many that I visit regularly, a few whose updates I try not to miss, and one that I check every day to learn about the most important web developments in our field. ...

This month, I picked two blogs to praise, and I owe readers some explanation. Now that everyone and their pets are blogging, you may experience blogorrhea. You need to be as selective in consuming blogs as in consuming meat and fruit in very slowly developing countries. There are many good blogging librarians and many excellent RSS feed manager programs on a very large variety of topics. Although I am still not an unconditional fan of blogs, because of the very high noise ratios, there are many that I visit regularly, a few whose updates I try not to miss, and one that I check every day to learn about the most important web developments in our field: ResourceShelf ([www.resourceshelf.com](http://www.resourceshelf.com)), created and managed by founding editor Gary Price, with contributing editor Shirl Kennedy.

Sometimes, I bump into a blog or a blog entry that shows something fascinating, such as the “listen to” feature in the blog of the Dutch blogger Guus van den Brekel (<http://digicmb.blogspot.com>), who was my fellow keynoter at the INFORUM 2007 ([www.inforum.cz/en](http://www.inforum.cz/en)) conference. He is using Odiogo, a Firefox audio plug-in, for converting text to audio. The sound is so good I did not want to believe my ears. Such discoveries make casual visits to blogs worthwhile.

After that detour, here is my column summary. The first pick is WoW!, the excellent blog about library and information science issues by the Dutch librarian Wouter Gerritsma. The WoW! blog is one I don't want to miss when it is updated (typically two to three times a month) with important developments in the fields of Wouter's LIS specializations.

The other pick is a service from Finland, the Feed Navigator of the National Library of Health Sciences, called Terkko. It aggregates feeds from thousands of well-selected medical resources and then organizes and presents the results very efficiently. It is a bonus that it also offers table of contents feeds from more than 50 library and information science journals.

The pan is IngentaConnect, a content aggregator that still does not offer full-text searching in the thousands of journals that it hosts. This is the exact opposite of what the best players in this genre (HighWire Press and MetaPress) and almost all the publishers of the scholarly journals in the Western Hemisphere do on their sites. It's also what the company's slogan promises: “Maximizing the value of content for publishers and libraries.”



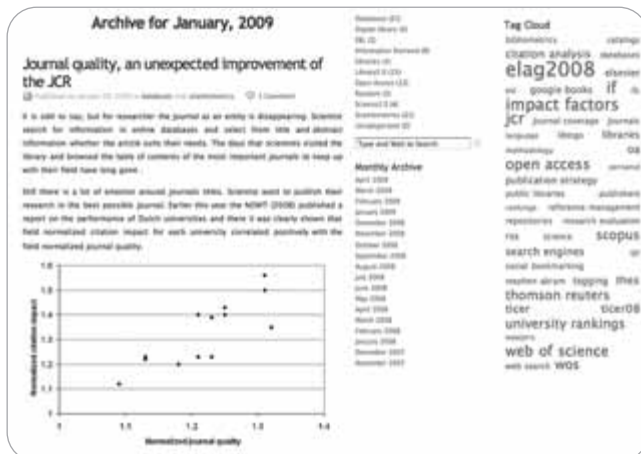
## the picks

### WOW!

WoW! ([www.wowter.net](http://www.wowter.net) in English; [www.wowter.nl/blog](http://www.wowter.nl/blog) in Dutch) is an acronym of the full title of this excellent blog site: Wouter on the Web (Wouter over het Web). Of course, I can't pass up the opportunity to say that it WoWed me a few years ago, and it keeps doing so, not by its size (he has only two to three entries a month in English) but by its timely and important content.

The blogger is a librarian at the Wageningen University in the Netherlands; he is an agricultural information specialist. This does not mean that his blog is agriculture-centric. Not at all. As you can see from the word cloud of his blog, he addresses several LIS issues. Many of these are of interest not only to our profession but also to the wider community concerned with bibliometrics, scientometrics, and journal impact factors. Research results in these areas have significant implications for collection development policies and actions as well as users' reactions to cancellations and non-subscriptions. I would certainly demote the listing of the links to the monthly sections of the archive to give visual priority to the text cloud (which includes some unnecessary and redundant abbreviations, such as IF for Impact Factor and OA for Open Access. I'd prefer putting them on a stop-word list if the blogware allows it; some independent cloud-generators do).

The topics very much coincide with my main interests, so I can rather objectively judge the topical choices and opinions: I like both. It sits well with me that he talks (blogs) when he has something to say that is important and of instant utility in these topic areas. I can also very well empathize with his occasional imperfections in using English as a second language.



Snapshot from the WoW! blog showing tag cloud

The blog entries approach issues pragmatically, reflect having done good homework, and feature well-presented observations without navel-gazing philosophizing—not the unusual ivory tower approach. The key points are well-illustrated by small but still legible screen shots within the flow of the commentary, and there are good links to pages for background. The entries are not in-depth reviews, but they also aren't just regurgitations of press releases. They are the right length for calling attention to a new or revised database or smart/dumb software features and choices.

The blog is also up-to-date. For example, I learned from this blog in January 2009 about a new edition of the *Journal Citation Reports* from Thomson Reuters, which was suspiciously unusual, as the new edition is always released midyear. It turned out that this was a revised release of the existing edition due to the addition of highly important new features. These features include a 5-year impact factor and two additional performance indicators for nearly 8,000 journals, using the eigenfactor model. In calculating the scholarly journals' impact factor, the algorithm weighs the number of citations received by a journal by incorporating the impact factor of the citing journals in the computation.

### TERKKO FEED NAVIGATOR

Terkko ([www.terkko.helsinki.fi/feednavigator](http://www.terkko.helsinki.fi/feednavigator)) comes from the National Library of Health Sciences in Finland. Among its many worthy services, there is a very important free one. It is called the FeedNavigator, which organizes and presents good quality RSS feeds. It is certainly not the marquee line on the homepage, but it seems to be very practical even if you are not a medical librarian.

This stood out for me because it focuses on journal tables of contents and feeds of widely respected nonspecialist media (BBC, CNN, and blogs); it does so superbly—no muss, no fuss.

It offers the best way for medical librarians to keep their patrons happy with reliable and relevant information about the most current articles of thousands of medical journals, health-related news in the media, and in the heavily filtered blogosphere. As a bonus, it does the same for quite a number of library and information science and technology sources, so nonmedical librarians can also benefit.

Terkko's biggest advantage for all librarians is that they can prepare themselves for an avalanche of questions when a popular TV station, radio program, or news magazine breaks important health news, often referring to some recently published research papers in medical or multidisciplinary journals. Patrons don't care whether there is a medical reference desk in your library or if you are a medical librarian. They want to get information from you about an issue. After all, in their eyes, you are *the* librarian at any time of the day, in any wing, section, or floor of the library.

It is gratifying to see how fresh the feeds are—and the designers make a point of it, by indicating the amount of time that has passed since the feed was received, which is often just minutes before you open up the site. It is a very

smart idea to be able to filter the scope of the feeds displayed to journals, news, and blogs by clicking on the filter line. For medical professionals, the first category is likely to be the most important; for the general public, it's the second and the third ones. There is a fourth category, for Finnish language sources. Although I am fluent in a Finno-Ugrian language, it is not Finnish, so I can't comment on that segment.

There are 4,323 information sources in FeedNavigator. It is true that there is a larger, free table of contents service, the excellent ticTOCs ([www.tictocs.ac.uk](http://www.tictocs.ac.uk)), a JISC-funded project. (The ticTOCs Consortium consists of the University of Liverpool Library [lead], Heriot-Watt University, CrossRef, ProQuest, Emerald, RefWorks, MIMAS, Cranfield University, Institute of Physics, SAGE Publishers, Inderscience Publishers, DOAJ [Directory of Open Access Journals], Open J-Gate, and Intute.) It centrally provides feeds from the tables of contents of 14,000-plus scholarly journals in all disciplines. FeedNavigator has many non-journal sources (with full-text open access content). Because of the significant differences in disciplinary and genre coverage, the number of sources and feeds should not be compared.

Terkko's sources are classified under 75 categories (mostly broad health topics, plus a few others), including a special category, Library and Information Sciences, not limited to the medical librarianship angle. Feeds are received from more than 50 LIS sources (depending on how we count the different sections of *EContent*), including *Journal of Documentation*, *GeoInformatica*, *Internet Research*, *JMLA*, *Library Hi Tech*, *Online Information Review*, and Information Today, Inc.'s NewsBreaks).

While I was enjoying a break (and my favorite red meat dish) during the testing of this site, a news item caught my eyes and made me rather unhappy, as a not exactly vegetarian civilian. It was about the consequences of red meat consumption. According to a longitudinal study, this habit is much more lethal than I knew. A follow-up search made the situation worse for me but better for FeedNavigator, since it showed off FeedNavigator's awesomely wide and relevant source base.



Excerpt from FeedNavigator's results list for a search on red meat

Navigation, browsing, linking, and personalization are a blissful software experience. I am not as happy with the search functionality, although I appreciate the fact that the sources can be searched by title, abstract, and author, which is not an option in ticTOCs, at least not yet. I strongly disagree with the decision to consider a space between two words as an OR operator. This was the approach in generic search engines a decade ago, and, not surprisingly, it turned out to offer results lists with extremely low relevance. It should be changed to an AND operator, as has happened in all the major search engines.

I also would like to see the number of hits that come with a search, so I can decide if I need to refine the search rather than clicking like crazy on the Next button to keep going ... and going. Currently, Terkko FeedNavigator uses IngentaConnect for searching journals, which is not the best choice for a reason I'll explain in the Pans section.

These developers should be able to craft their own federated search engine for the publishers' sites they are using for feeds. Some years ago I tried to do something like this—without hits consolidation—for 15 of the largest publisher sites. It was one of my most satisfying hobby projects, except for the fact that publishers kept changing their sites' structures and search syntaxes, which would have required constant modifications in my polysearch engines, so I could not continue it. The much more sophisticated developers at Terkko could and should go on enhancing this excellent service by extending the search options for these sources to the full text of the journals. It is certainly not an expectation from a table of contents service, even in the digital stratosphere (as opposed to a digital journal hosting service). But it would make users happier, and it would enrich their experiences tremendously with this top-notch service.



## the pan

### INGENTACONNECT

Times they are a-changin' indeed. In 1990, Ingenta was my pick in this column with its free and novel services. Even then I was concerned that the full text of the journals was not searchable, which has become a standard feature on almost all the best hosts' (digital facilitators) and publishers' sites, with the notable exception of the hosting services by AIP Scitation (which was my pan earlier) and IngentaConnect ([www.ingentaconnect.com](http://www.ingentaconnect.com)).

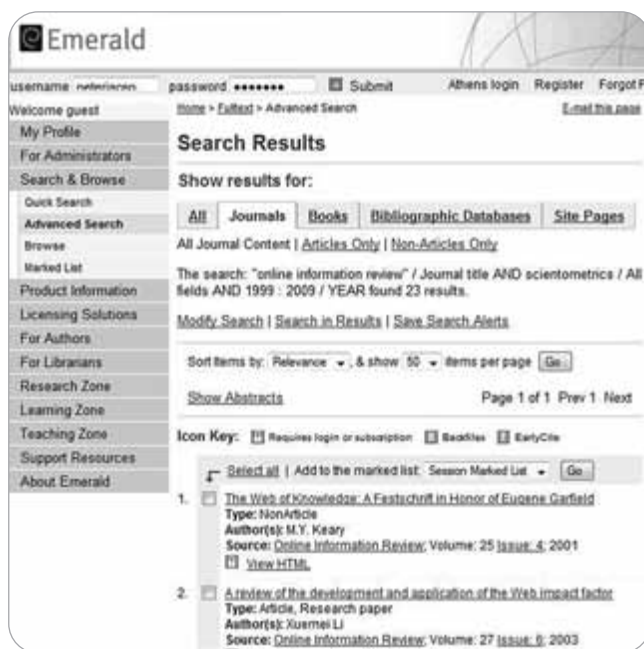
Let me make it clear that I don't expect the full text to be free—just the searching of the full text for better (far better) resource discovery. When Ingenta acquired my then favorite CatchWord hosting service—which did offer free

full-text searching of the journals it hosted—I hoped that the acquisition would not make this and other excellent features disappear. I was told not to worry. Well, they did disappear, and for almost a decade, there were only minimal enhancements in the IngentaConnect software.

In spring 2006, I brought up this issue again at the excellent UK Serials Group ([www.uksg.org](http://www.uksg.org)) conference at the University of Warwick when talking with the company's very competent representative. I was told that full-text searching is coming. As of the end of 1Q 2009, it still isn't available.

Full-text searching makes a huge difference in recall (and is likely to cause some loss in precision—depending on the type of information needed and the query formulation). For example, a search for the term *blogging* in the full text of papers published in *Online Information Review* between 1999 and 2009 would find 13 hits when using (for free) its publisher, Emerald's, search engine ([www.emeraldinsight.com](http://www.emeraldinsight.com)). In IngentaConnect, the same search in the title, abstract, and keyword fields produces only two hits. It is true that the search term may only be mentioned in passing in some of the full-text hits. The search term *scientometrics* produces 23 hits through the native search engine, while in IngentaConnect, the number of hits is 0.

Of course, one could test and reveal this only if the journal(s) is available on IngentaConnect and from the publisher. This is the exception rather than the rule. I chose *Online Information Review* because I am familiar with its content and software and because—luckily—it does not rely on IngentaConnect. This has serious implications for searchers and also for the publishers that have journals that are hosted only by IngentaConnect. Searchers using IngentaConnect may wrongly assume that *Online Infor-*



Search results for *scientometrics* using Emerald

*mation Review* does not publish any papers that discuss issues about or related to scientometrics.

It does not help that on Ingenta there are records for only 354 items from the same number of journal issues (56) of *Online Information Review*, while on the Emerald site this number is 838. The huge difference is caused by IngentaConnect failing to include the book reviews or the substantial entries in the Literature Guide section—this is quite a distortion and a disservice for the users, the authors, and the publisher.

Of course, I appreciate that there is a site that offers searching for free on the title, abstract, and keyword fields of 26,086,612 articles, chapters, reports, and other materials in 32,087 publications. But in 2009, this is like giving away PCs with 360KB of floppy drives and 256KB of RAM to developing countries. I am not expecting the equivalent of an OLPC (One Laptop Per Child) model, but this author-title-keyword searching does not make sense.

More importantly, the slogan of the company suggests that its managers don't realize this when proudly claiming "Maximizing the value of content for publishers and libraries." This seriously handicapped search module does not maximize the value of content for either the publishers or the libraries, let alone for the researchers. It does the opposite and has a negative impact on the services that rely on IngentaConnect.



Search results for *scientometrics* in IngentaConnect's title, abstract, and keyword fields

*Péter Jacsó* ([jacso@hawaii.edu](mailto:jacso@hawaii.edu)) is professor of library & information science at the University of Hawaii's department of information and computer sciences.

Comments? Email the editor ([marydee@xmission.com](mailto:marydee@xmission.com)).

