

## Web indexing results 99

**Note: many of the links are no longer current.**

Last year, I took over the management of the Web Indexing award from Dwight Walker. In consultation with the National Committee, I initiated some changes in the rules; the main one being that 'gateway' sites, where one Web page links to many other pages scattered throughout the Web, would no longer be eligible for entry. We took this approach largely because: a) the skills involved in producing and maintaining a gateway are not necessarily those involved in indexing; and b) I had no way of judging what constituted a 'good' gateway site for a particular topic without having a specialised knowledge of the topic and the Web sites relating to it.

Whether because of this change or due to limited marketing, there were only three entries for the 1999 AusSI Web Indexing Prize:

- The Case In Point Index at <http://www.acxiom.com/caseinpoint/cip-ix-home.asp>  
This was last year's runner-up, and was submitted by Dave Ream and created by Acxiom.
- Bowne Internet Solutions Index  
This was submitted by Cheryl Lemmons.
- The Pre-Raphaelite Critic Index at [http://www.engl.duq.edu/servus/PR\\_Critic/Reviews.htm](http://www.engl.duq.edu/servus/PR_Critic/Reviews.htm)  
This was submitted (and compiled) by Thomas J. Tobin.

Rather than try and pick a winner from this small field, the National Committee decided to send each entrant a subscription to the AusSI Newsletter and I agreed to review the indexes here. Links to these indexes can be found at the AusSI free-for-all links site at [links.html](#). My thanks to Belinda Weaver, of inCite, who volunteered to help judge the indexes when we were still uncertain how many we would get.

### **Case In Point**

The Case In Point index is maintained by Acxiom to provide users with access to Articles, Case Studies, and Reports stored on-line. Most of these are marketing reports. There are several indexes with different entry points. Each index has a set of alphabetical links at the top to take the user down to the section of the page where that letter starts, but there is no quick way to return to the top of the page.

There is no introductory matter in the index and no attempt to explain or guide users to the correct entries. No mention of or credit to the indexer(s) is given, and there is no obvious way for the user to provide feedback on the index.

Acxiom customers are displayed in bold, and the type of article is shown in the index.

Throughout all the indexes, references take the user to the beginning of the article rather than the point at which the mention is made, meaning that the user has to scroll down or do a text search to locate the material. Readability could be improved by adding margins, and navigation could be improved by putting 'Return to Top' buttons at regular intervals (or using frames).

The different indexes were:

- Associations and Organizations – there were only seven entries here and it seemed rather pointless to separate them off from companies; of the seven references one produced a 'Page Not Found' error.
- Companies – this was an index of mentions, rather than of information about the companies referenced; for instance, a reference to 'Compaq' led to an article containing a one-line quote from a Compaq executive talking about the computer industry in general. There's no way to distinguish these 'mentions' from articles in which a company is discussed at length. The Companies index also includes entries like Forrester Report, which would have been better placed under Subject.

- People – this is an alphabetical list of people mentioned or quoted in the articles. The same problems arise as with Companies in distinguishing between mentions and more detailed information.
- Products – these are Acxiom products mentioned in the articles.
- Subjects – the Subject index involves genuine content analysis, including See and See Also cross-references and subheadings; e.g.

Data consolidation/integration

in credit-card industry

CS: Cross-Industry Case Briefs

in insurance industry

CS: Cross-Industry Case Briefs

Internet methods

N5.4: "CRM with Networked Data"

outsourcing

N6.1: "Data Consultants"

It occupies about 600 lines and provides users with access to material of use to them. There are a few slips – ironically, one reference, given as N5.1 'Of Golf and Grandchildren' actually leads to an article called *Tracking Down Addressing Problems*. In general, though, Case In Point seems thorough and well-thought-out.

### **BowneInternet Solutions**

Bowne is a consulting company – that much seems clear, although the details of what it does and how are concealed behind a layer of tech-speak: "At Bowne Internet, we will work with you to develop an online strategy anchored in your corporate objectives, build your e-business from design through launch, and help you take your solution to the next phase". Er, yes, well... The index is to the site, which appears to contain about fifty pages.

This index begins with a graphics-intensive page that slows downloading, but once received the layout is excellent. The margins and the return-to-top links that are absent from the Case In Point Index are used to good effect here, and the shortness of the index (about 150 lines) means that it can be spread out without becoming too tiresome to move through. The index also contains links to text within pages as well as to the top of a page, making it easier to locate the material you are after. Two levels of sub-headings are used in one location (under 'Success Stories'); only one level of subheading is used elsewhere. The choice of colours is excellent and makes it easy to read.

Most of the items in the index are the titles of pages on the site; as far as I can tell there has been no attempt at subject analysis apart from providing See cross-references. Where the page titles are obscure an explanatory gloss is provided in parentheses: e.g. Evolution (Bowne Internet service to clients).

There are many cross-references, but as each cross-reference is itself a link to the destination page, it would have been simpler and less confusing to omit the see altogether and just go straight to the destination page: e.g. 'Approach - see How We Work' actually links, not to the 'How We Work' line in the index, but to the Web page entitled 'How We Work'. As it is, the number of See references makes the index look a little off-putting and inadequate.

In summary, the Bowne Internet Solutions index is attractively laid out but contains relatively little subject analysis; it is more or less an alphabetical list of the titles of pages on the site. Again, there is no credit or acknowledgement for the indexer and no explanatory matter at the head of the index describing what it is for and how it works.

### **The Pre-Raphaelite Critic**

This is a one-man production, compiled by Thomas Tobin to provide details of Pre-Raphaelite criticism (ie. art and literary criticism of and by Pre-Raphaelites), and access to online examples where they exist. It is part of Tobin's attempt to find and place online all examples of the genre.

There are three indexes (although the plural is given as *indices*) – by author, publication date, and periodical title – and there is some explanatory matter at the head of each index.

The indexes are not as useful as they might be because they are laid out as large (400kb) tables which extend beyond the boundaries of my (1024-pixel) screen and require scrolling across, unless set to the smallest text size; a paragraph layout which allowed for margins would have made the data easier to read and possibly reduced the size taken up by so many table-related HTML tags. Breaking the tables up over several pages – A-C, D-F – would reduce the download time. A set of alphabetical navigation buttons would also help. There is no subject analysis here – just an alphabetical listing of authors and titles and a chronological listing of publication date.

- Author index – Here anonymous authors are filed as blanks and listed at the top of the author sequence, Since these make up nearly half the articles, this means that the ‘alphabetical listing’ of authors defaults to a chronological sequence for these items. They are followed by a set of tentatively identified authors in square brackets, meaning that if you want to look up articles by, say, Oscar Wilde, you have to look in two places: under ‘Wilde, Oscar’ and under ‘[Wilde, Oscar]’.
- Chronological index – this has been updated recently and the table has lost its intrusive borders, although the layout is still confusing.
- Periodical title index – probably the most useful index for general purposes.

These tables are not so much an index as the raw material for an index; non-scholarly users will find them too large to browse through, and the Edit/Find option is mainly of use when looking for particular authors, dates or periodicals. A more generally useful approach would be to prepare a combined Subject Index listing both authors and periodical titles, combined with categories for the type of article – review, parody, etc – the work or works being discussed and their originators, and the subject-matter. For instance, the work currently indexed as

1851 05 17 20 00 219 *Punch Punch among the Painters* 75.01  
might generate the following index entries:

*Alisma Plantago* (Collins)

Collins

*Alisma Plantago* (painting)

Millais

*Mariana* (painting)

painting

errors in proportion

proportion

errors in painting

*Punch* (magazine)

*Punch among the Painters*

*Punch among the Painters* (1851)

Obviously we are looking at a considerable amount of work here (any volunteers?).

### **Conclusion**

Each of these indexes represents a great deal of work, and they all have some attractive features, but they could all be improved. Some basic principles for Web indexing can be identified which any indexer should be able to implement.

### **Things to include:**

- Introductory material, stating what this is an index to, how it works, when it was updated, who made the index and how to get in touch with them. This should also indicate the size of the index and its depth.
- Alphabetical navigation buttons to jump through the index and ‘Go to Top’ buttons to return.
- Links to internal destinations within pages as well as to pages themselves.
- Legible layout with margins and a readable choice of colours
- See and See also references (where appropriate) to other items in the index.

### **Things to avoid:**

- Multiple indexes – these have been shown to confuse people using books, and there is no reason to think the Web is any different.

- Cross-references that go to the material on site.
- Bulky graphics (or other files) that delay downloads.
- Overlong pages not broken into alphabetical sections.