

An e-Interview with Ali Shiri

Author of *Powering Search*

Dr. Ali Shiri is an associate professor in the University of Alberta's School of Library and Information Studies. He teaches a course in advanced topics in the organization of knowledge, including indexing, abstracting, and thesaurus construction. He is the author of Powering Search: The Role of Thesauri in New Information Environments (Medford, NJ: Information Today, 2012). Moira Calder interviewed him by telephone.

MC: Information professionals and researchers, including indexers, come from a wide variety of backgrounds and work in many areas.

AS: Indeed. ... The role of the indexer is more diversified now and more versatile, I believe. When we talk about indexers, we're obviously talking about different types of indexes. It can be back-of-the-book indexes or indexes created by databases vendors, providers, or corporate agencies. Now, this whole area of digitization, the mass digitization projects around—you see a lot of content is actually digitized and is available online—in fact provides a wider range of opportunities for indexers to actually be involved in and to use their skills. I always say to my students in my indexing class that these skills are transferable, and if they know the general principles of indexing, it is a set of very powerful techniques that they can use for the analysis and representation of information.

As for the indexer, in my book I'm basically trying to capture these new information environments using an umbrella term to refer to a wide range of digital libraries, institutional repositories, content management systems, open archives. They are in some form or fashion new information environments. But you could even label them as new information indexing, representation, and retrieval environments. I think the value and the quality of indexing obviously play an important role in this whole process because we are still talking about how we represent the subject about-ness of items. And with all these new search engines, even with Google being the number one search engine for the past 10 years, there are a lot of digital repositories that basically apply metadata and indexing approaches simply because general free-text searching is inherently problematic. It may be useful for certain search tasks but not for all tasks ...

MC: What about e-books, specifically Amazon's advice to its authors not to include an index?

AS: I must say that I was not aware of that statement on Amazon. I think that's a very radical approach, because I can name two specific projects, one in the U.K. and one in Canada. These are research projects, funded projects by SSHRC [Social Sciences and Humanities Research Council] and one in the U.K. They actually focus on the notion of meta-indexes. What that means is, basically, how can we best benefit from the back-of-the-book indexes in a digital environment? One of the projects looks at how we can take advantage of back-of-the-book indexes in the electronic books to create a superindex, using that subject structure, that intellectual asset ... to create a knowledge map or a map of a particular domain or different domains. If you are talking about e-books in the area of civil rights or human rights, for instance, you can actually create a very nice index using back-of-the-book indexes for that domain and incorporate access points from different sources into one big index. What I'm saying is to actually advocate for the importance of these ... indexes.

Another project in the U.K.—the title of the project is BOBI, Back-of-the-Book Indexes—on usability ... examined the search and interaction behaviour of 45 students looking at three different electronic books. They designed search tasks and compared three different finding and retrieval tools. One was the back-of-the-book index, one was free text, and the third one was the table of contents. The result of that study was that definitely the back-of-the-book indexes for those electronic books outperformed the two other approaches, followed by the table of contents. That means that this structure that BOB indexes provide certainly adds value to the whole search process as users interact with the electronic book [see references about this project below].

Now I haven't talked yet about thesauri ... The one at UBC, at the University of British Columbia, talks about the notion of a meta-index [see references about this project below] for mapping the terminology of a domain using back-of-the-book indexes. These two projects, I believe, are trying to leverage the power of indexes in a way, at least in terms of back-of-the-book indexes in electronic books.

MC: What do you think indexers need to know about other information knowledge domains?

AS: In the book that I published, part of my argument was that there are certainly a number of domains that indexers can become more familiar with. Indexers have been involved in this process of subject analysis and indexing for decades. Now you will see these new tools and all of the digital [attempts at indexes]; some of them are actually reinventing the wheel. You'll see that they are struggling with what LIS professionals and indexers have been struggling with for decades, and they have even found solutions, but people think they're finding something new or different. Particular areas, like information architecture and information search behavior, I think, indexers need to become familiar with or at least read. I think that reading the research literature on the information search behaviour, information interaction behaviour, and information retrieval behavior of users helps them actually to re-find the value of their skills. You know, we could be talking about indexing in our own domain for hours and months and years, but the value of it and how it can be applied to relevant issues and questions—in other words, relevant search and retrieval issues that we are currently wrestling with—even Google wrestles with. So if I want to name those areas, it would be information architecture and information search behaviour.

Also, probably, the area of metadata. I see that some of these areas, basically, are converging, because if you look back a few years ago, even Google entered into a contract with WorldCat because they found the value of metadata and intellectual analysis of documents rather than an automatic analysis of content by thousands of crawlers. I think indexes and index tools are very similar. It's a matter of how we can actually apply this knowledge and skills that the indexers have to new information environments or environments that are different: not just books or e-books but subject databases, portals, subject repositories, digital repositories, open archives, as I mentioned. So to answer your question, I think it is important for them to read the literature on information search behaviour. I'm not talking about broad information behaviour or human information behaviour. People could be sources of information. I mainly talk about more interaction with computerized information services—search behavior, web search user studies. Those are very useful. I'm sure they will find things that they have answers for and that many people struggle with mainly because they don't look at them from an indexing or an information representation and retrieval purpose.

MC: Natural language versus controlled vocabulary?

AS: I know that there is an argument about natural language processing and applications. You know, the indexing of content has come a long way and has made progress. I remember, back in 2006 in Vancouver, I attended the Information Architecture Summit for the first time. There were about 600 web designers, information architects, webmasters, and managers. They were enthusiastically and excitedly talking about how they can use facet analysis to create nice organizational structures, information structures for websites, as if they had found something new, whereas facet analysis was introduced by Ranganathan back in the 1960s.

Why does this community of information architects think that some form of vocabulary control is important for findability and retrievability and searchability of content? At least for a look at current search systems and search tools, you don't see fully functional natural language processing-based tools that function perfectly. That's why a lot of corporate websites use indexing and metadata applications to provide more focused and more efficient access points ... When you look at the literature on controlled vocabulary versus free-text searching, there are a number of articles published over the past 30 years or so. The short, brief, and succinct outcome of all these studies is that these two even now complement each other. Each lends itself to different types of search tasks. That's what we can say for sure. That's why even corporate websites such as ebay or Zappos—online shopping sites—try to somehow provide predetermined or preset categories for faceted browsing purposes, so that users don't have to get lost [in searches] purely based on free-text searching and thinking up terms. They're probably not aware of what the system or the site is using or has used before.

These are some of the challenges that we are facing, I guess. Google and Ask.com are the best natural language processing tools around. They are good for certain tasks, right? But Google started introducing some level of metadata for refining and narrowing down searches by date, by time, even by document type. So it's very clear that users would like to have browsing functionalities and features, not just a simple search box.

MC: Any comments generally?

AS: Yes. I think what I can say, as I have said in the book, is that there are many thesauri and controlled vocabularies available that could nicely be used in new information environments and in indexing in particular. I have worked with some of the people who are actually working on a number of different metadata projects, and my students have. What they have said is that these skills are quite transferable in their capacities as metadata or metadata developers or people in charge of digital libraries or digital metadata. I think indexing has a

significant role in these new information environments. It's just becoming more familiar, first of all, with this wide range of information environments and then how these indexing skills can be introduced to those different environments. Finally, because thesauri tend to be domain-specific, they are going to be useful for content management systems and digital libraries and subject gateways. I think that those are relevant to the current search landscape and digital search on the Web.

Resources

Back-of-the-Book Index (BOBI) usability project:

Abdullah, N., and F. Gibb. "Using a Task-Based Approach in Evaluating the Usability of BOBIs in an E-book Environment." In *Advances in Information Retrieval*, 246–57. Berlin: Springer, 2008.

Noorhidawati, A., and F. Gibb. "How Students Use E-books: Reading or Referring?" *Malaysian Journal of Library and Information Science* 13, no. 2 (2008): 1–14.

Meta-index project:

Huggett, M., and E. Rasmussen. "Dynamic Online Views of Meta-indexes." In *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries*, 233–36. ACM, June 2012.

———. "Meta-index Views on Digital Domains." In *Proceedings of the 2012 iConference*, 618–19. ACM: February 2012.

Upcoming Conferences

The Australian and New Zealand Society of Indexers (ANZSI) conference was held in March 2013. The title: "Intrepid Indexing: Indexing Without Boundaries." Reports from this conference are on the society's website (http://www.anzsi.org/site/conference_papers.asp). This society meets every other year.

The ISC/SCI Annual Conference, "Between the Lines / Entre les lignes," will be held June 6–7, 2013, in Halifax, in conjunction with the Editors' Association of Canada conference. Stay up-to-date at <http://indexers.ca/annual-conference/>.

On July 12–14, 2013, the Society of Indexers (U.K.) will meet at Wyboston Lakes conference centre in Bedfordshire. The conference theme is "Revitalize Your Business, Refresh Your

Skills." Information is available at <http://www.indexers.org.uk>.

"Words in 3 Dimensions," a conference devoted to writing, publishing, and editing, will be held in Edmonton on May 24–26, 2013. On May 25 local ISC/SCI representatives will set up an information booth to promote our association and indexing in general. Information is available at <http://www.wordsin3d.com/>.

The American Society for Indexing 2014 conference will be held April 30–May 3 in Charleston, South Carolina. Information is on the ASI website (<http://www.asindexing.org>), under Conferences.



Heather Ebbs came across this magpie while in New Zealand representing ISC/SCI.

Book Review:

Powering Search: The Role of Thesauri in New Information Environments

By Ali Shiri

“The art and science of organizing and labeling ... to support usability and findability”: does this sound familiar to you as an indexer? Insert the words “websites, intranets, online communities, and software” in the space, and you have the definition of information architecture cited on page 15 of Ali Shiri’s book *Powering Search*. A theme that runs through this thought-provoking book is that many people in many disciplines are grappling with the same issues and that we can and should learn from each other. (See the interview with Dr. Shiri in this issue for more of his thoughts.)

When I raised the possibility of reviewing this book for the *Bulletin*, one reaction was “Will this interest indexers?” After all, many, if not most, create indexes for standalone books using natural language. True enough, but controlled vocabulary is an area that indexers will increasingly need to understand and address. EPUB 3 recommends that a “reading system must be able to link from one EPUB to another.”¹ Furthermore, at the 2012 ISC/SCI conference, Cheryl Landes described her vision of the future of indexing as including indexing databases and search optimization (see the *Bulletin*’s Summer/Fall 2012 issue). What role will indexers play in “leverage[ing] the power of linked data” (p. 130)?

Shiri provides an overview of thesaurus terminology, standards, and uses in digital environments as indexing tools and as adjuncts to search engines. He discusses research on controlled vocabularies as complementary to natural language searching (p. 242). Three chapters address interface design, and in the final chapter Shiri presents an overview of current research. On first read I found the parts on how people look for information the most interesting; on second look I found the design sections led me to

reflect on future interactive indexes. After reading the section on faceted search user interfaces (pp. 8–15), I started noticing online indexing where I hadn’t thought about it previously and thinking of the possibilities. For example, on the Chapters website I can search an artist under “House and Homes” if I want a poster or under “Books” if that’s what I want.

Powering Search has a well-written index that includes authors cited and subject headings and identifies figures and tables, which helped me to locate information that I wanted to reread. Because it is a scholarly publication, the reference lists at the ends of chapters are portals to more information unto themselves, and although some of the examples of databases provided are proprietary, you can access some online without a subscription, and Shiri provides many screen shots of interfaces. I also found the diagrams of information-seeking and retrieval models helpful. That being said, this book is not a light read. It was written for information scientists, not for indexers. The presentation is scholarly and its content is research focused.

One of my first “big” freelance contracts some years ago involved an index for an online database of health publications. My tech colleagues had never worked with an indexer before and asked me to make a presentation. The first question I fielded: “What do you do?” I might have asked them the same question. Shiri’s book will help facilitate that dialogue among information workers.

Moira Calder

Note:

1. International Digital Publishing Forum. *EPUB 3 Indexes Charter*. <http://idpf.org/charters/2012/indexes/>