

Reviews

Darin L. Stewart. *Building Enterprise Taxonomies*. Mokita Press, 2008.

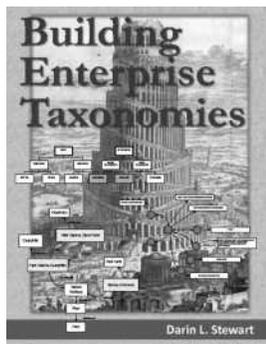
Controlled vocabularies, thesauri, and taxonomies are closely related to indexing, especially periodical and database-type indexing.

More recently interest in the field has spread from the world of library and information science to corporate content

management, knowledge management, and website information architecture. Thus, the latest books on taxonomies tend to be aimed at the latter applications. The newest book, *Building Enterprise Taxonomies* by Darin L. Stewart, is one of these. While the book does not deal with indexing, indexers who would like to know more about taxonomies in general might still be interested in this book, as there are few books dedicated to taxonomies.

Actually, this book presents human indexing as something undesirable. Stewart lists among the possible goals for a taxonomy, the "elimination of manual indexing" (p. 84). Stewart also states: "One of the most common applications of enterprise taxonomies currently is support of website navigation" (p. 53), but mere navigation does not necessarily require indexing. Even manual taxonomy work is presented as undesirable. The author advises: "Wherever possible, avail yourself of automatic data gathering tools. Unfortunately, true taxonomy work will always boil down to hands-on, human driven consideration and labor" (pp. 82-83). Why, I wonder, is that unfortunate? We taxonomists want work! Apparently Stewart's audience is not necessarily professional taxonomists.

If you come across *Building Enterprise Taxonomies* during a search on books on taxonomies, as I did, you might wonder who is Darin Stewart, what is Mokita Press, and what is the intended audience of this book. Even after you obtain the book, you might still wonder the same. The book lacks an introduction or preface to explain its purpose or audience and to put it into context, and Mokita Press has no Web presence. In fact, a negative review of the book on Amazon.com derided the book: "Quite frankly, I could not figure out who's the target audience for this book." (A. Goldberg) It turns out that the book is self-published and the fictitious publisher name was created for this book.



Stewart, who holds a Ph.D. in information science (2000), is the director of Web Strategies and Information Services within the Information Technology group of Oregon Health & Science University, and he also teaches a continuing education course in the Applied Information Management Master's Degree Program at the University of Oregon. As he explained in an e-mail to me, he wrote this book for the course he teaches. (The origin of the book in a course slips through on p. 68: "As pointed out in the first lecture...") This understanding of the audience and background of the book is important for the reader's expectations.

The first chapter, "Findability," starts out explaining how difficult and time-consuming it is to find desired information in the glut of information there is on the Web and why search engines are inadequate. Numerous statistics are provided, and studies on search behavior are also cited. The chapter concludes by saying that controlled vocabularies and taxonomies can provide the solution. "A taxonomy can act as a dynamic bookmark" (p. 22).

The second chapter is devoted to the subject of metadata. A basic introduction to metadata, types, and standards is provided, but the connection to taxonomy is not made strongly enough. In fact, taxonomy is never even mentioned in the metadata chapter, except for the final sentence. Instead, "authority control" is mentioned as what is needed to make metadata initiatives successful, and taxonomies are said to be the "lynchpin," without further explanation as to how.

Taxonomies, not introduced until Chapter 3, are described as a means of organizing and classifying. Stewart defines taxonomy as "a hierarchical collection of categories used to organize information," and that taxonomies are a kind of controlled vocabulary, which is more complex than an authority file but less complex than a thesaurus. Thus, taxonomy seems to be defined narrowly. The chapter entitled "Taxonomy," however, includes sections on controlled vocabularies.

"Enterprise taxonomy," however, is never defined. It becomes clear, especially in the fourth chapter, "Preparations," that enterprise taxonomies are used to organize all the information within a company or organization and not necessarily for human indexing. It is explained that the content to be covered with the taxonomy may include invoices, purchase orders, compliance documentation, personnel action forms, meeting minutes, procedural documentation, correspondence, and standard operating procedures (pp. 90-91).

Terms are the focus of the fifth chapter. The discussion on sources for terms, both within an enter-

Submissions and suggestions for reviews should be submitted to the Editor, keywords@asindexing.org.

prise and externally, is quite useful. The explanations of term format follow the ANSI/NISO Z39.19 standard quite closely. In fact, some of the examples were taken from the standard. The explanation on format and style of terms seems to get a little muddled with scope notes given as examples for named entities, and the statement that named entity terms are better kept out of a taxonomy and in an authority file instead. I found some inconsistent information here, where the author states that “it is best to use a non-trade name as an entry term where possible” where “entry term” clearly means term or preferred term, but earlier in the book “entry term” is defined as a nonpreferred term.

After discussing terms and their formats, the sixth chapter is about structure. Facets and hierarchical relationships are both adequately discussed. I find that a full explanation of the various types of permitted broader/narrower relationships is crucial for the understanding of taxonomies, and I am pleased that Stewart gives that topic sufficient attention. The explanation of the associative relationships, however, is relatively weak.

Chapter 7 takes up an important topic, interoperability, as the software tools used to create taxonomies are usually not those used for searching for content with a taxonomy. This chapter discusses XML, XSLT (Extensible Stylesheet Language Transformation, although Stewart omits definition of the T), and Zthes. While the explanations could be clearer, the inclusion of these topics is significant. According to the author in a list post, “I believe this is the only book currently available that specifically discusses the application of XML-based technologies and standards (including XSLT, RDF, Zthes, SKOS, etc.) to managed vocabularies” (Taxonomy Community of Practice, May 13, 2008). Part of the reason why the explanations on how XML is used with taxonomies are a little confusing is that not enough space is given. To treat the subject sufficiently, however, would probably require two whole chapters instead of only part of one. XSLT, something with which I and probably most readers are not familiar, also lacked clear explanation. The section on Zthes I felt needed more background information on what this is, who or what organization is behind it, and to what extent it is followed.

Towards the end of the book is a chapter on ontologies. A lengthy explanation of ontologies in a book focused on taxonomies is welcome, as ontologies are often discussed more from the computer science, rather than the library/information science, realm. However, the book jumps from simple hierarchical taxonomies to the complex ontologies without giving much discussion to thesauri with a mid-level degree of complexity in their relation-

ships and term attributes. There is in fact a continuum of complexity among various kinds of taxonomies and thesauri. Indeed there are other books that more adequately cover thesauri. The ontology chapter also covers the Resource Description Framework (RDF) and Web Ontology Language (OWL) data models. I found the treatment of these topics more theoretical than practical and again not as clear as I would have liked. I was pleased, however, that Stewart mentioned the leading open source software tools for creating ontologies. Unfortunately, he does not mention any taxonomy software. Is there a taboo on mentioning commercial software in a book not dedicated to software?

The final chapter is on folksonomies, now a standard related topic to taxonomies. Here Stewart describes tagging and tag clouds and presents the examples of *flickr.com* and *librarything.com*, among others. The question left unanswered is how this applies specifically to *enterprise* taxonomies. I have heard that social tagging can be effective with a critical mass of contributors to public websites, but may not work inside an organization. While the connection of folksonomies to taxonomies is worth mentioning, it probably does not require a full chapter in this book.

Since I am already familiar with taxonomies and terms, it's a little difficult for me to evaluate whether Stewart's explanations in these areas are sufficiently clear to the novice. Where I have less experience is in project planning and management. I did find his advice useful on how, as the taxonomist, to prepare for one's initial meeting with the taxonomy project sponsor. Stewart gets rather detailed in recommendations such as a 90-second initial talk with the sponsor followed by a 90-minute meeting with the initial stakeholders. He explains the components of a governance document, such as a statement of purpose in 30 words or less and a signature page with a line for each member of the steering committee (p. 86). These suggestions and those of milestones are indeed useful for the new independent taxonomy consultant. I also found his instructions on conducting a content audit helpful. Additional business knowledge seems to be assumed, however. In addressing the creation of a governance document in preparation for a taxonomy project, Stewart states: “If possible you should provide a business case to illustrate and justify these decisions” (p. 86). I would have appreciated it if Stewart could elaborate on possible business cases.

I do not fully agree with all the advice given in this book. For example, Stewart says that you first gather and properly format 100-500 terms and then you start putting them in a structure. In practice, you can start building

the structure at the same time that you gather the terms. He also explains that the method to put terms into a structure is by having stakeholders perform card-sorting exercises and that, by dividing up the cards of terms, as many as 500 terms can be sorted by several groups of people. Although I am not an expert in enterprise taxonomy construction, I still feel inclined to question this approach as being the only one. Card-sorting exercises make lots of sense for small website navigational taxonomies or for questionable parts of a taxonomy hierarchy, but there is a lot of variation in size and scope of enterprise taxonomies, so different approaches might be needed. Stewart also explains that one first creates relationships among terms and then creates categories, and then the categories are organized into an overall structure. This is essentially a bottom-up approach. It is generally accepted that there are two possible methods for creating taxonomies, bottom-up or top-down.

Stewart's assertion that “associative relationships do not necessarily need to be reciprocal” (p. 134) is false to my understanding of taxonomy standards. According to the ANSI/NISO Z39.19-2005 guidelines, “A basic property of relationships in controlled vocabularies is that they are reciprocal.” This contrasts with cross-references in book indexes, which might be what was on Stewart's mind. It is permissible to have a *See also* reference without a reciprocal one. But in thesauri or other controlled vocabularies the RT (related term) relationship should exist at both terms of the relationship pair. In fact, all thesaurus maintenance software packages always enforce this reciprocity. It sounds like Stewart is confusing *reciprocity* of relationships with *symmetry* of relationships, when he further states: “It is perfectly valid to include one-way or asymmetric relationships” (p. 134). Relationships between terms are always reciprocal, but not always symmetrical, as in the case of broader term/narrower term, which is an asymmetric reciprocal pair of relationships.

Stewart also makes several unreferenced generalizations that I would question. For example: “When done properly, metadata initiatives nearly always generate a positive return on investment. Unfortunately, few are done properly and most fail” (pp. 41-42). Personal opinions are stated as fact, such as: “Term selection is also the most difficult aspect of creating a taxonomy” (p. 89). In other cases, perhaps a citation is needed for statement, such as: “S. R. Ranganathan, generally considered the greatest librarian of the twentieth century...” (p. 60), and “Despite the explosion of the taxonomy software and services industry of the past several years (it has been growing at an average annual rate of 21% since 2002) the

vast majority of taxonomies are still created and maintained in Microsoft Excel” (p. 137). I’d love to know that source of that 21% statistic, and I also disagree that most taxonomies are created and maintained in Excel, based on my own survey of taxonomists.

While there appear to be a fair number of diagrams, illustrations, or graphics, some may not be entirely appropriate and others lack sufficient explanation, beyond a short caption. For example, a schematic diagram on page 148 with arrows between elements (content management system, external taxonomy, other channels, and enterprise search engine) has a caption “the secret life of an enterprise taxonomy” and no explanation in the text on that or adjacent pages, which are discussing XML instead. Is the caption supposed to be cute? Other graphics are web page screen shots, which have so much busy content and are reproduced rather small, it leaves me wondering what I’m supposed to be looking for within the screen shot image, as the text does not describe the image, and the short caption does not explain anything other than “a web page with good information scent” or “the advanced search page of the LA Times.” Graphics quality is otherwise adequate, especially considering that the book is self-published.

While taxonomy and term examples have been chosen for clarity and simplicity, for a book on enterprise taxonomies I would have found it more instructive to have examples common in the business or corporate world. Instead, most examples are from biology or food and wine. Even an example of an ontology for people deals with students rather than employees/experts.

Various terms are bolded—such as taxonomy, enterprise taxonomy, controlled vocabulary, entry terms, user warrant, etc.—but there is no explanation as to why they are bolded. The book contains no glossary. If this book is in fact used as a class textbook, then perhaps these are terms that the student is required to know.

There are many typos and formatting glitches, which may be typical of a self-published book. For example, on page 4, it says parenthetically, “See sidebar on opposite page,” but there is no sidebar on the opposite page. In an example of three scope notes, one ends with a comma, the next one ends with a period, and the third ends with no punctuation. In other places punctuation is missing, and on a number of pages the text switches in and out of black type to gray type at odd places for no apparent reason. In the endnotes, author names are sometimes first name – last name and sometimes inverted. Furthermore, some of the URL links listed, such as for

existing taxonomies, are dead. For a book less than a year out of publication, I would expect better for the website links.

I didn’t find anything really original in this book. So, I guess it works well as a textbook then, since a textbook presents established information. In the final chapter, Stewart depends too much on other people’s words, though, as the four pages 206-209 contain about an equal amount of text in a number of indented quotes as the author’s own body of text.

Finally, the index of the book is perhaps the worst I’ve ever seen. It has no subentries and dozens of undifferentiated locators at many of the headings. I’ve seen this before, but in this case the strings of page numbers are especially long since page ranges aren’t even used. Cross-references are not used correctly, as entries have both a *See* reference and a locator. The referenced headings had the same locator in one case and different locators in other case, so there wasn’t even any consistent pattern to the incorrect use of *See* references. Double posts are not used correctly. For example, the two entries Variant terms and Nonpreferred terms (which I would consider to be the same thing) had different locators. There is an entry for Abstract concepts but no entry for Concepts. Topics of discussion were often missing from the index, such as any entry for names/named entities/proper nouns. Indexed topics were not fully covered. For example, scope notes were sufficiently discussed on page 104, but in the index the locators are for pages 57 and 111 only. Style is also inconsistent. We find index entries in the plural for: Candidate terms, Non-preferred terms, Preferred terms, and Related terms; but in the singular for Broader term, Narrower term, Precoordinated term, and Top term. It’s not surprising that a self-published book does not have a professionally written index. What is surprising is that an expert on taxonomies, even if not trained on indexing, could have done such a poor job on an index. He had even mentioned the need for consistency in plural form for countable nouns as taxonomy terms (p. 106), but he could not apply that in his index.

Despite its faults, the book has some merits. I’m not sure if I would recommend it to someone completely new to taxonomies, because there are some inaccuracies and generalizations. But for anyone who already has a good understanding of taxonomies, controlled vocabularies, or thesauri, and would like to know more about creating enterprise taxonomies in particular, this book definitely has some good tips and thus would be of value. This book appears to more than adequately fulfill its purposes as a course textbook, although I would have liked to see less than a

full chapter devoted to findability and folksonomies and other areas expanded instead. Except for the omission of commercial software, *Building Enterprise Taxonomies* brings together, although a bit disjointedly, all the topics of interest on the subject.

— Heather Hedden,
Manager of the ASI
Taxonomies & Controlled
Vocabularies SIG

Information and Emotion: The Emergent Affective Paradigm in Information Behavior Research and Theory, Diane Nahl and Dania Bilal, eds. Information Today, 2007.

In *Information and Emotion*, Diane Nahl and Dania Bilal bring together a number of peer-reviewed studies in the field of information behavior (IB) research. While this volume is largely aimed at the information scientist, there are several articles of interest to the indexing community, especially Web indexers. As Nahl explains in the introduction, the volume “...demonstrates the deepening realization in information science that technology is deployed within a mental environment of emotions and thoughts” and that there exist both “information environments” and “information communities” reflecting the social and communal aspects of information technology (p. ix).

The collection addresses these issues on a variety of levels, beginning by addressing theoretical concerns and then moving on to examine the “macro-informational information environment” and the “micro-components of emotionality that shape the information experience and condition” (xx). Finally, the last section documents specific cases of “emotional intensification of the information environment when a break occurs in human symbiosis with technology, due to physical disability or demographic estrangement from the information culture” (xx).

Part I: Theoretical Frameworks provides a useful introduction to the approaches underlying IB research. Most useful, particularly to the neophyte, in this section are Nahl’s “The Centrality of the Affective in Information Behavior,” in which Nahl lays out the theoretical foundations of IB studies and then discusses the two phases of interaction with informa-

