

The accidental taxonomist

Heather Hedden

Information Today, Inc.

Medford, New Jersey

2010

xxix + 442 pp.

ISBN 978-1-57387-397-0

\$39.95

George Macgregor

Information Management & Systems

Liverpool John Moores University

Keywords information retrieval, taxonomies, controlled vocabularies

Although taxonomic classifications have been an important strand of information organisation and retrieval for years, the word 'taxonomy' has become particularly omnipresent in information, computer and library science literature over the past decade. Writing in 2000, Alan Gilchrist noted the sudden rise of corporate taxonomies as a response to issues of information overload and the need to better organise corporate digital information assets (Gilchrist, 2001). Indeed, author of *The accidental taxonomist*, Heather Hedden, charts the rise of the 'taxonomy' buzzword in Chapter one, noting the exponential growth in literature on the aforementioned topic from 2000 onwards, with a huge spike in 2005. Interest in taxonomies remains high in 2010, although the choice of buzzword may have diversified to include folksonomies, ontologies, information architecture, and so forth.

Whilst taxonomies (and controlled vocabularies generally) continue to be deployed within digital libraries, information services and digital repositories, the principal growth area since 2000 has been within enterprise content management (ECM). It is ECM – and its variants – which has fuelled interest in taxonomies and it is ECM that, in a way, provides the motivation behind *The accidental taxonomist*. In chapter two ('Who are taxonomists?'), Hedden describes the emergence of the 'accidental taxonomist'; the IT systems officer who presides over an enterprise content management system but suddenly finds themselves appointed to create and implement an enterprise wide taxonomy, or the data entry clerk that is picked to create terms for a taxonomy because no one else more qualified was available. This sort of scenario might sound implausible, but it is one which appears to be borne out by Hedden's own research (published as an appendix in the book). Her survey of practicing taxonomists in late 2008 found only 48% to have a formal qualification in information science or library science. It is also interesting to note that Hedden has accrued almost 15 years experience in controlled vocabulary management and thesaurus construction (Hedden, 2010), but she is herself an accidental taxonomist having started her career as a writer.

With the agenda of the book encapsulated in its title, the book seeks to provide coal face guidance for practicing and novice taxonomists. The opening chapter delineates taxonomies and provides an overview of their role in indexing and information retrieval. It is worth noting that Hedden's definition of taxonomy is extremely holistic, encompassing not just taxonomies as they are generally understood, but also other knowledge organisation systems (KOS) (e.g. thesauri, subject heading lists, etc.). After a chapter dedicated to outlining the competencies required to become a taxonomist, Hedden provides chapters on term identification and creation (Chapter 3) and between term relationships (Chapter 4). Chapters pertaining to taxonomic structures (e.g. hierarchies, facets, etc.) and taxonomy interface and display issues are interrupted by chapters on taxonomy management software, manual indexing and automatic indexing which - according to Hedden's introduction - was an attempt to move towards "what you need to know ... in order of importance" (p. xxvii). Chapters

eleven and twelve provide useful guidance on taxonomy project planning, management, implementation and evolution. This also encompasses guidance on taxonomy interoperability, including discussion of eXtensible Markup Language (XML), applications of XML (i.e. Zthes) and the Resource Description Framework (RDF). The final chapter surveys the taxonomy profession and the nature of the job opportunities available.

Although the book provides wide and detailed coverage of many topics germane to controlled vocabularies, it is perhaps weakest in its treatment of interoperability issues, technical formats and RDF. In fact, on this latter topic, a mere three paragraphs are provided, with only a brief mention of the Simple Knowledge Organisation System (SKOS) and the Web Ontology Language (OWL) provided. ECM may have driven the creation of taxonomies over the past decade; but it is the Semantic Web which will drive their creation into the next decade. Nobody is expecting an authoritative guide to Semantic Web technologies, but such light treatment is a little perplexing. In the opening and closing chapters, Hedden makes a series of curious remarks which are also worthy of comment. The following is indicative:

“There is no undergraduate major or graduate degree in taxonomy and no department, program, concentration, or certificate in the field. Thus, people do not choose to be taxonomists when they decide what they want to study. Furthermore, the majority of graduate schools and programs of information science, or library and information science, do not have even a single course devoted to creating taxonomies (although it is often a topic within a course).” (p. xxiv)

Taxonomy courses may not exist and the word ‘taxonomy’ may not feature in module titles, but controlled vocabularies and their construction is an integral component of any graduate programme on information science or librarianship. Certainly, within the UK such topics are often subsumed by modules on information retrieval, knowledge organisation, classification and indexing, and so forth. Within my department, such topics are given detailed treatment in ‘information organisation’ and ‘information architecture’. Hedden appears to advocate dedicated taxonomist qualifications but it unclear how useful this would be in practice. Information scientists tend to have a suite of skills that can be deployed in a variety of information scenarios, and for good reason. These skills can be combined to solve particular information problems and it seems strange to restrict one’s skill set in these uncertain economic times by enslaving oneself to taxonomy. This is by no means a slight on the book or the author but rather a passing observation.

The accidental taxonomist is highly accessible and is particularly well suited to the novice taxonomist. Hedden’s writing style is easy going and difficult concepts for beginners, such as taxonomic, partonomic and instantive relationships, are capably deconstructed using numerous annotated examples. She makes reference to prevailing national and international standards where appropriate too, but thankfully stops short of bombarding the reader with the associated technical intricacies by prioritising things relative to the book’s intended audience. Screen dumps and diagrams also help to make Hedden’s coverage of taxonomy management software and interface design particularly strong. The book includes an appendix documenting the findings of Hedden’s taxonomist survey, as well as appendices for further reading, useful websites and a glossary. A highly detailed index is included, making the book extremely useful as a practitioner’s reference manual.

The accidental taxonomist manages to cover a wide range of controlled vocabulary topics at an intermediate level successfully. Let us be clear though, *The accidental taxonomist* is not for the experienced practitioner; rather, it is – as the title suggests – for someone that finds themselves constructing controlled vocabularies by accident, or someone who is relatively new to the topic area. Students at information studies departments the world over will doubtless find Hedden’s accessible

writing style and well selected examples refreshing too. Still, hardened professionals may find its coverage of taxonomy project management makes it useful as a reference book or as an emergency refresher text. In some ways *The accidental taxonomist* does itself a disservice by conflating a variety of different KOS under the term 'taxonomy'. The book is far more than just taxonomy. Indeed, its title is such that readers interested in, say, thesauri will not necessarily consider it a worthwhile purchase unless they study its contents closely. This tragedy is almost ironical given the book's intellectual content and could easily have been obviated had an edifying strap line been included.

References

Gilchrist, A. (2001), Corporate taxonomies: report on a survey of current practice, *Online Information Review*, Vol.25 No. 2, pp.94-102.

Hedden, H. (2010), Heather Hedden's resume, *Hedden Information Management*, available: <http://www.hedden-information.com/resume.htm> (accessed 28 July 2010)