

SECOND EDITION

The  
Accidental  
Taxonomist



Heather Hedden

SECOND EDITION

The  
*Accidental*  
Taxonomist



Heather Hedden



Information Today, Inc.

Medford, New Jersey

*First Printing, 2016*

***The Accidental Taxonomist, Second Edition***

Copyright © 2016 by Heather Hedden

All rights reserved. No part of this book may be reproduced in any form or by any electronic or mechanical means, including information storage and retrieval systems, without permission in writing from the publisher, except by a reviewer, who may quote brief passages in a review. Published by Information Today, Inc., 143 Old Marlton Pike, Medford, New Jersey 08055.

**Publisher's Note:** The author and publisher have taken care in preparation of this book but make no expressed or implied warranty of any kind and assume no responsibility for errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of the use of the information or programs contained herein.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book and Information Today, Inc., was aware of a trademark claim, the designations have been printed with initial capital letters.

**Library of Congress Cataloging-in-Publication Data**

Names: Hedden, Heather.

Title: The accidental taxonomist / Heather Hedden.

Description: Second edition. | Medford, New Jersey : Information Today, Inc., [2016] | Includes bibliographical references and index.

Identifiers: LCCN 2016002968 | ISBN 9781573875288

Subjects: LCSH: Information organization. | Classification. | Indexing. | Subject headings. | Cross references (Information retrieval) | Thesauri.

Classification: LCC Z666.5 .H43 2016 | DDC 025—dc23

LC record available at <http://lcn.loc.gov/2016002968>

Printed and bound in the United States of America

President and CEO: Thomas H. Hogan, Sr.

Editor-in-Chief and Publisher: John B. Bryans

Associate Editor: Beverly M. Michaels

Production Manager: Tiffany Chamenko

Marketing Coordinator: Rob Colding

Indexer: Kathleen Rocheleau

Cover Designer: Ashlee Caruolo

Composition by Amnet Systems

**[www.infotoday.com](http://www.infotoday.com)**

# Contents

<b>Figures and Tables</b> .....	vii
<b>Foreword</b> , by Joseph Busch .....	xi
<b>Acknowledgments</b> .....	xv
<b>About the Website</b> .....	xvii
<b>Preface to the Second Edition</b> .....	xix
<b>Introduction</b> .....	xxiii
<b>Chapter 1: What Are Taxonomies?</b> .....	1
Definitions and Types of Taxonomies .....	1
Taxonomies and Metadata .....	15
Applications and Purposes of Taxonomies .....	18
Taxonomies for License .....	28
History of Taxonomies .....	32
<b>Chapter 2: Who Are Taxonomists?</b> .....	43
Backgrounds of Taxonomists .....	44
Taxonomist Skills .....	57
Related Duties .....	60
Employment of Taxonomists .....	62
<b>Chapter 3: Creating Terms</b> .....	77
Concepts and Terms .....	77
Identifying Concepts .....	80
Choosing the Preferred Term .....	87
Term Format .....	93
Precoordinated Terms .....	95
Notes and Attributes .....	99
<b>Chapter 4: Creating Relationships</b> .....	105
Equivalence Relationships and Nonpreferred Terms ....	106
Hierarchical Relationships .....	118
Associative Relationships .....	126

## **iv The Accidental Taxonomist**

Hierarchical/Associative Ambiguities .....	132
Semantic Variations for Relationships .....	134
<b>Chapter 5: Software for Taxonomy Creation and Management .....</b>	<b>141</b>
Software Not Designed for Creating Taxonomies .....	143
Dedicated Taxonomy Management Software .....	149
Single-User Desktop Thesaurus Software .....	159
Multiuser Taxonomy Management Software .....	162
Free and Open Source Software .....	178
Other Software with Taxonomy Management Components .....	183
<b>Chapter 6: Taxonomies for Human Indexing .....</b>	<b>193</b>
What Is Human Indexing? .....	193
Terms, Relationships, and Notes for Indexers .....	200
Taxonomy Structure and Indexing Interface .....	204
Taxonomy Updates and Quality Control .....	209
Managing Folksonomies .....	215
<b>Chapter 7: Taxonomies for Automated Indexing .....</b>	<b>221</b>
Automated Indexing, Search, and Taxonomies .....	221
Automated Indexing Technologies .....	226
Software for Auto-Categorization .....	239
Creating Taxonomies for Automated Indexing .....	251
<b>Chapter 8: Taxonomy Structures .....</b>	<b>257</b>
Hierarchies .....	257
Facets .....	270
Multiple Vocabularies and Categories .....	277
<b>Chapter 9: Taxonomy Displays .....</b>	<b>283</b>
Thesaurus Displays and Outputs .....	284
Hierarchical Taxonomy Displays .....	294
Fielded Search and Faceted Displays .....	305

<b>Chapter 10: Taxonomy Planning, Design, and Creation</b> .....	315
Planning for a Taxonomy .....	316
Enterprise Taxonomies .....	329
Taxonomy Creation Process .....	336
Taxonomy Governance .....	344
<b>Chapter 11: Taxonomy Implementation and Evolution</b> .....	349
Taxonomy Exchange and Interoperability .....	349
Taxonomy Updating .....	357
Combining Taxonomies .....	361
Multilingual Taxonomies .....	375
<b>Chapter 12: Taxonomy Work and the Profession</b> .....	383
The Nature of Taxonomy Work .....	383
Taxonomists as Contractors .....	391
Education and Training .....	398
Organizations, Networking, and Resources .....	407
<b>Appendix A: Survey of Taxonomists</b> .....	413
<b>Appendix B: Glossary</b> .....	423
<b>Appendix C: Recommended Reading</b> .....	441
<b>Appendix D: Websites</b> .....	447
<b>About the Author</b> .....	455
<b>Index</b> .....	457



# Figures and Tables

Figure 1.1	Terms in a synonym ring . . . . .	5
Figure 1.2	Terms in an expandable hierarchical taxonomy . . . . .	7
Figure 1.3	A term in the <i>ASIS&amp;T Thesaurus</i> . . . . .	11
Figure 1.4	A domain ontology dealing with retail management . . . . .	14
Figure 1.5	Library of Congress Subject Headings . . . . .	21
Figure 1.6	Medical Subject Headings . . . . .	22
Figure 1.7	Top-level taxonomy of Amazon.com . . . . .	24
Figure 1.8	Faceted taxonomy of the Microbial Life Education Resources search site . . . . .	25
Figure 1.9	The Information Architecture Institute site map, a navigational taxonomy . . . . .	26
Figure 1.10	Example taxonomy offered through Taxonomy Warehouse . . . . .	30
Table 1.1	Number of periodical articles about taxonomies . . . . .	39
Figure 3.1	A spreadsheet setup for a content audit, with a unique row for each concept . . . . .	82
Figure 3.2	A spreadsheet setup for a content audit, with a unique row for each source file or document . . . . .	82
Table 4.1	Types of nonpreferred terms . . . . .	111
Figure 4.1	Hierarchical relationship characteristics . . . . .	119
Figure 4.2	Polyhierarchy for the term <i>light trucks</i> . . . . .	124
Figure 4.3	Polyhierarchy for the term <i>Great Salt Lake</i> . . . . .	125
Figure 4.4	Associative relationship between terms in different hierarchies . . . . .	127
Table 4.2	Types of associative relationships . . . . .	128
Figure 4.5	Associative relationship between sibling terms with overlapping meaning . . . . .	129
Figure 4.6	No associative relationship between sibling terms with no overlapping meaning . . . . .	130
Table 5.1	Survey responses for types of software used . . . . .	143
Figure 5.1	Using Excel to create a five-level taxonomy . . . . .	145
Figure 5.2	Using Excel to create a taxonomy for import . . . . .	146



**viii The Accidental Taxonomist**

Figure 5.3	Mind map display in FreeMind .....	148
Figure 5.4	MultiTes taxonomy display with selected term details .....	161
Figure 5.5	Data Harmony Thesaurus Master's split-screen display .....	165
Figure 5.6	PoolParty Thesaurus Server's split-screen display .....	168
Figure 5.7	Semaphore Ontology Editor's split-screen display .....	171
Figure 5.8	Synaptica's hierarchical taxonomy display and selected term details .....	174
Figure 5.9	TopBraid Enterprise Vocabulary Net user interface .....	176
Figure 5.10	Microsoft SharePoint's Term Store .....	189
Figure 7.1	Synonym ring supporting search for a concept .....	225
Figure 7.2	WebStudio user interface .....	245
Figure 8.1	Two different methods of displaying a geographic hierarchy .....	259
Figure 8.2	Facet indicators for types of automobiles .....	269
Figure 8.3	Facets for men's shirts on the Sears retail website .....	276
Figure 9.1	Simple list alphabetical display .....	286
Figure 9.2	Flat-format alphabetical arrangement of terms .....	287
Figure 9.3	Full-term hierarchy displays .....	290
Figure 9.4	Top term hierarchy display excerpts .....	291
Figure 9.5	ERIC Thesaurus's A-Z browse display .....	293
Figure 9.6	USDA Thesaurus's search options for terms ...	294
Figure 9.7	One level per webpage from the Dex Media Superpages .....	296
Figure 9.8	Expandable tree hierarchy of the <i>USA Today</i> Content Tree .....	299
Figure 9.9	Fly-out subcategories on the site Lynda.com .....	301
Figure 9.10	Hierarchy of book categories on Amazon.com .....	304
Figure 9.11	Facets for limiting product reviews on the Buzzillions website .....	307

Figure 9.12	Facets for library materials at the North Carolina State University Libraries . . . . .	308
Figure 9.13	Fielded search options for clinical trials at ClinicalTrials.gov . . . . .	311
Figure 9.14	Advanced Search of Gale Business Collection database . . . . .	312
Figure 10.1	OptimalSort card-sorting software . . . . .	342
Figure 11.1	XML output extract from MultiTes . . . . .	351
Figure 11.2	ZThes output extract from the NASA Thesaurus . . . . .	352
Figure 11.3	OWL output extract from the NASA Thesaurus . . . . .	354
Figure 11.4	SKOS output extract from the NASA Thesaurus . . . . .	355
Table 11.1	Automated term matches that require review by a taxonomist . . . . .	368
Figure 11.5	Mapping table example . . . . .	372
Figure 11.6	VocBench screenshot with Linked Open Data URLs . . . . .	374
Figure 11.7	Multilingual concept in the AGROVOC Thesaurus . . . . .	376
Figure 11.8	SKOS XML of a multilingual concept . . . . .	381



# Foreword

Organizing electronic content using metadata fields with controlled vocabularies has at least a 50-year history. It's the story of how we got from expensive, rarely used time-shared databases to the almost ubiquitous web where anyone can "look it up" anywhere, anytime. The work of tagging content has always been done by an army of indexers, more geeks than librarians, working in more of a cottage industry than a factory. All were accidental information scientists with backgrounds in business, medicine, law, the humanities, and maybe sometimes library but rarely computer science.

Some people may think that the content in Heather Hedden's practical compendium is "old wine in a new bottle," but somebody had to write this stuff down. True, librarians have been doing cataloging, classification, and subject indexing for a long time, long before electronic content became a format to manage. But meaningfully adapting appropriate practices from library science and communicating them in a form that can be effectively used by a broad interdisciplinary audience is the major accomplishment of this book.

Taxonomies to support content indexing and finding could be tied to the history of database systems that included processable text information. At first these databases were electronic versions of abstracting and indexing services offered as very expensive, time-share online services (e.g., Dialog), later as subscription CD-ROM databases, and most recently as various types of web-mediated services. In the early days, two disciplines dominated the online services—medicine and law. Medical informatics was heavily subsidized by governments (especially in the United States) after World War II, and legal information (e.g., LexisNexis) was valuable enough to be paid for by large corporations who were the clients of large law firms. Medical Subject Headings (MeSH) was introduced by the National Library of Medicine in 1960. Its precursor was the

## **xii The Accidental Taxonomist**

subject headings of *Index Medicus*, which date from 1940. Medical “subjects” are one of the taxonomy gold standards. They include taxonomies for the human body, taxonomies for conditions and treatments, taxonomies for medical practice settings, etc.

The iterations of digital environments over the past 50 years have had major impacts on what would be considered effective and efficient information organization strategies. In the era of expensive, time-share online services, taxonomies needed to enable especially precise retrieval because every minute and every citation to an information source had a significant cost associated with it. End users, such as business managers, were typically not allowed to execute their own searches. This was an era of intermediated searching. The online searcher (often a librarian) was a highly trained gatekeeper and often a subject matter expert him- or herself.

With CD-ROMs the costs of online access were eliminated. But the content organization schemes had to be changed to work on these self-contained platforms. The web changed this again, at first replacing content organization with the power of web search engines (Google, Yahoo!, Altavista, etc.); global taxonomies, such as the DMOZ Open Directory Project; and very importantly, online shopping. Search engines transformed us into a “look it up” culture. Shopping online has taught everyone how to do Boolean searching, these days referred to as search refining.

The current era of the semantic web is proving to be a further watershed, because its underpinnings are the identification of named entities—people, organizations, locations, events, products, topics, and the like—when they occur in the content on the web. The first-generation web enabled the observation and boosting of content relevance based simply on its access and use. The semantic web is enabling the identification of relationships among all types of named entities and the presentation of information based on these relationships. Simply put, the semantic web is based on the organizing power of faceted taxonomy.

Inside the organization, the relatively new current expectation is that information should be as findable and linkable as on the public web. Enterprise applications are more and more becoming web services that happen to be within the organizational firewall. Employees expect there to be

- a single place for internal information delivery
- a view of information across different business silos
- easy access to others across different business groups to foster collaboration
- a trusted location for conducting day-to-day activities

As taxonomy becomes a ubiquitous part of the organizational information ecosystem, there is more and more demand from organizations for people who have the skills to integrate taxonomies into enterprise applications. But what exactly does creating and maintaining taxonomies entail, and where are you going to find the appropriate expertise to competently undertake these tasks? While this is a great time to be a taxonomy consultant, one measure of the success of one of our engagements is whether a taxonomy editor has been identified or hired to be the central point of contact for taxonomy maintenance. Hence, you may find yourself becoming an “accidental taxonomist.”

This book is an excellent primer for the novice who finds him- or herself assigned (or volunteering for) the task of creating and maintaining a taxonomy. The book should also serve as a “bible” for the expert (I have a copy on my shelf). It answers these key questions I am frequently asked:

- What is a taxonomy?
- Who are taxonomists?
- How do you create, maintain, and use a taxonomy?
- Where can you find taxonomy tools?

## **xiv The Accidental Taxonomist**

This edition is a comprehensive revision, notably updating screenshots of websites, renewing the section on taxonomy software, and adding information about two important new taxonomy standards: ISO 25964 (*Thesauri and Interoperability with Other Vocabularies*) and SKOS (*Simple Knowledge Organization System*), a W3C recommendation.

As a consultant, I am a proponent of keeping things as simple as possible. *The Accidental Taxonomist* is a very useful resource for me to share with my clients and prospects. It is full of information about the various considerations related to content organization and is one of the best sources for guidance on best practices for addressing them.

—Joseph Busch

*Joseph Busch is an authority in the field of information science and a frequent speaker at conferences on metadata and taxonomy. Prior to founding Taxonomy Strategies, a consulting firm that guides organizations in improving information capture, preservation, search, retrieval, and governance, he was vice president for Infoware at Metacode Technologies and the Getty Trust's program manager for Standards and Research Databases. He is a past president of the Association for Information Science and Technology (ASIS&T) and a member of the Dublin Core Metadata Initiative Executive Committee.*

# Acknowledgments

The field of taxonomies is broad and multidisciplined, and thus no single taxonomist-author could comprehensively cover the subject without getting input and insights from others with different experiences. I am very grateful to the colleagues of mine who have contributed to this book. I especially want to acknowledge the very thorough review of the first edition of my book by two expert reviewers, Lynda Moulton and Alice Redmond-Neal, who voluntarily took time out of their busy schedules to meticulously go through every chapter. In addition to providing suggestions for better wordings and corrections to any inaccurate generalizations and assumptions, they each contributed some additional bits of information and insights from their own experience that I incorporated. Additional expert review on certain sections and chapters came from Margot Diltz and Tom Reamy.

I want to give a big thank you to my friend and indexing colleague Jean Jesensky for giving my book its initial read-through for clarity. Thanks also go to the taxonomy consultants I interviewed and the other taxonomists who completed my online surveys. For the second edition, I am grateful to Enterprise Knowledge for letting me log into their SharePoint 2013 account so that I could experiment with the Term Store and take a screenshot.

This book would not have been possible without the hard work of the staff of Information Today, Inc., especially John Bryans, who encouraged me to write both the first and second editions. John and his team helped make the first edition a success, leading to seven printings, and ably supported my many book signings in the years that followed its release.

In addition to those who helped directly with this book, I wish to acknowledge all those who contributed indirectly, such as by teaching or mentoring me or by giving me opportunities. My primary



mentors in the area of controlled vocabularies and taxonomies were my manager at what was then Information Access Company (now Cengage Learning), Margot Diltz, and consultant Jessica Milstead. Through Gale's partnership with Synaptica, I also received feedback on my early vocabulary work from Trish Yancey. A later mentor was Joseph Busch, from whom I learned more about taxonomy implementation, including taxonomy testing and governance. I also appreciate the demo access that various taxonomy software vendors set up for me to experiment with the various products. Even the students I taught in my online courses helped me indirectly by asking questions, which I realized needed to be addressed. So I also thank Kris Liberman for giving me the opportunity to teach numerous sessions of my taxonomies workshop through the continuing education program of Simmons College School of Library and Information Science.

Finally, thanks go to my husband, Tom, also a freelancer, who had to work more to help financially support us during the year I wrote the first edition, as I was taking on less consulting work, and also for proofreading selected passages of my second edition.

# **About the Website**

## **[www.accidental-taxonomist.com](http://www.accidental-taxonomist.com)**

*The Accidental Taxonomist* mentions a number of websites throughout. These include examples of online taxonomies, software vendor sites, and useful reference sites. These websites are all listed in Appendix D. To facilitate quick access to these resources, a website with links to all these related sites can be found at [www.accidental-taxonomist.com](http://www.accidental-taxonomist.com). Furthermore, the links will be maintained and updated on the website as needed over time, and new sources may be added. Please email your comments, changes, and suggested additions to the author at [heather@hedden.net](mailto:heather@hedden.net).

### **Disclaimer**

Neither the publisher nor the author makes any claim as to the results that may be obtained through the use of this webpage or of any of the internet resources it references or links to. Neither the publisher nor the author will be held liable for any results, or lack thereof, obtained by the use of this page or any of its links; for any third-party charges; or for any hardware, software, or other problems that may occur as the result of using it. This webpage is subject to change or discontinuation without notice at the discretion of the publisher and author.



# Preface to the Second Edition

When I published the first edition of *The Accidental Taxonomist*, I knew that changes would be needed within a couple of years, mostly to reflect the changes in thesaurus management software vendors, as software is a volatile industry characterized by new companies, acquisitions, and some vendors going out of business. It was also expected that the website examples, given as screenshots in the book, would change. As it turned out, the changes were more widespread than anticipated. I ended up replacing all screenshots and adding some new ones (totaling 44), since even existing software vendors or websites had updated their user interfaces. More than half of the various website URLs found throughout the book also had to be updated.

In the area of software, what I did not anticipate was that software changes have gone beyond just who the vendors are and what features vendors have added. There have also been some notable trends, such as in the adoption of Semantic Web standards, the convergence of taxonomy and ontology support, and more web-based, cloud/software-as-a-service offerings. Thus, in addition to adding more software vendors (and removing a few), I have also added a short section summarizing all of these software trends.

Also with respect to software, the first edition made no mention of SharePoint, since SharePoint 2010, the first version to support taxonomies, came out the same year my book did. So this new edition includes some discussion of managing taxonomies in SharePoint. There is not the space here to go into all the details, so I explore specific topics, such as managing polyhierarchy in SharePoint, on my blog, also called *The Accidental Taxonomist*.

## **xx The Accidental Taxonomist**

The standards have changed too. ANSI/NISO Z39.19 2005 *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* was reaffirmed in 2010, but more significantly ISO 2788 *Guidelines for the Establishment and Development of Monolingual Thesauri* and 5964 *Guidelines for the Establishment and Development of Multilingual Thesauri* have been replaced by ISO 25964 *Thesauri and Interoperability with Other Vocabularies*, Part 1 in 2011 and Part 2 in 2013. This is not merely a reorganization of parts. The changes also comprise new content in the area of interoperability, including the exchange of taxonomy data and mappings between vocabularies. Now ANSI/NISO Z39.19 is coming due for a new version, but it is a long process. With an eye to a wider international audience, in this edition I cite the ISO standard along with the ANSI/NISO standard whenever relevant.

In addition to the change in the ISO thesaurus standard, there is also a change involving the wider adoption of other kinds of standards, most significantly those associated with the Semantic Web. Although development had begun earlier, the World Wide Web Consortium (W3C) formally released the SKOS (Simple Knowledge Organization System) standard only in August 2009, when I was busy finalizing my manuscript for the first edition, before the extent of the eventual adoption of SKOS was still unknown. Now it is quite common for taxonomy management software to follow the SKOS specifications of concept modeling and taxonomy output. So, more attention to SKOS is given in this edition.

Another trend, which was already underway at the time I wrote my first edition, but which I simply did not bother to consider in detail, is the convergence of metadata and taxonomy. So, I have added a short section on the topic. I needed the intervening years to actually work in areas where taxonomies and metadata meet, whether through consulting or in a department called Metadata Standards and Services, before I felt I could say something original on the subject.

As for the people who do taxonomy work, the accidental taxonomists, I conducted a new survey, which has shown that their backgrounds remain as diverse as they were when surveyed six years prior, but there are new stories and examples of how people got involved in this type of work and what they like about it. Meanwhile, the opportunities for taxonomists continue to grow. I executed the exact same search for jobs in fall of 2009 and again in fall of 2015, on the job board aggregator Indeed.com, and found the numbers of currently posted openings had significantly increased.

Although I considered myself quite experienced with various taxonomies at the time I wrote the first edition, I have continued to gain additional taxonomy work experience since, so here and there throughout the book I have added information based on further reflection. Thus, in the chapter on planning and designing a taxonomy, I have added some advice regarding designating facets for enterprise taxonomies, questions to ask during stakeholder interviews, how to conduct stakeholder workshops, and methods of testing taxonomies

I had also started writing my blog the year after the first edition, but the blog post topics are not the same as the additions to this book. The Accidental Taxonomist blog allows me to explore tangents in more detail, and this book is already longer than needs to be!

Taxonomies are interesting in that some things about them are fundamental and do not change, such as the notion of a concept, its varied names, its hierarchical and nonhierarchical relationships with other concepts. But, as anything related to information technology, there are things about taxonomies that do change, such as how they are managed, implemented, and utilized. Thus, it is not only the varied subject matter that makes taxonomy work interesting, but also the various implementations and opportunities to take advantage of in new technologies, such as those related to the Semantic Web and Linked Open Data. Although this new edition addresses these topics, my ongoing blog will cover further considerations in such areas.



# Introduction

After reading a case study of an enterprise taxonomy in which corporate research librarians were charged with the task of building the taxonomy,<sup>1</sup> it occurred to me that many people who get involved in creating taxonomies do so by accident. Even if this case study is not typical, it illustrates the point: The growing interest in taxonomies means that the people being asked to create taxonomies may not have done that work before, may not have sufficient training, and/or may not even have thought of pursuing such work before they were asked to. This hypothesis was borne out by responses to an online questionnaire I wrote, in which taxonomists explained how they got into the field.

Most of us first became familiar with the term *taxonomy* in high-school biology when the concept was used in reference to the classification and naming of plants and animals. If you did not pursue a career in biology, you probably did not give the concept any further thought for quite some time after that. Although the term is also used to refer to nomenclature and classification of concepts in other academic disciplines, only since the late 1990s has it been understood to mean information organization in general. Taxonomy in this sense includes controlled vocabularies for document indexing and retrieval, subject categories in content management systems, navigation labels and categories in website information architecture, and standardized terminology within a corporate knowledge base. In some of these areas, such as websites, the application of taxonomy is relatively new, coinciding with the newer adoption of the term *taxonomy*. Other areas, such as controlled vocabularies and thesauri used in periodical indexing and literature retrieval, have been around for decades. Their publishers may continue to refer to a “controlled vocabulary,” an “authority file,” or a “thesaurus,” even though the newer usage of the term *taxonomy* is also used for these purposes.



Today there are many meanings of the word *taxonomy*, which can complicate any research into the term. Although the original meaning, the *study* of classification, is rarely used, the term *taxonomy* continues to be used to designate classification systems of things. Originally used for the classification of things in nature, the term spread from the sciences to the social sciences and thus came to be used also for the classification of concepts. (One better-known example of such taxonomies is the Taxonomy of Educational Objectives, also known as Bloom's Taxonomy.) Despite the recent popularity of the term *taxonomy* for generic knowledge organization, the majority of books and scholarly articles on taxonomies in print today are still about highly specific classification systems in the sciences or social sciences. The taxonomists of those systems are experts in their academic disciplines rather than librarians or information architects.

Even as a generic system of knowledge organization, the term *taxonomy* presently has two different common usages. One meaning of taxonomy, reflecting the earlier usage for the classification of living organisms, is a hierarchical classification of things or concepts in what may be considered a tree structure. Terms within the taxonomy each have a "parent," or broader term, and a "child," or narrower term, unless the terms are at the very top or bottom levels of the taxonomy. Another, even more recent, usage of the term *taxonomy* is to refer to any controlled vocabulary of terms for a subject area domain or a specific purpose. The terms may or may not be arranged in a hierarchy, and they may or may not have even more complex relationships between each other. Thus the term *taxonomy* has taken on a broader meaning that encompasses all of the following: specific subject glossaries, controlled vocabularies, information thesauri, and ontologies. Each of these will be explained in further detail in Chapter 1. For the purposes of this book, this second, broader definition of taxonomy is used. It is the simplest term, and it corresponds to the word taxonomist.

As the word taxonomy has different meanings, so does the designation of a taxonomist. It can still refer to a biologist who specializes in the field of naming and classifying organisms. The majority of people with the title of “taxonomist” today, however, are information specialists, librarians, or information architects and are not likely to be subject matter experts. They deal with taxonomies in the broader definition of knowledge organization systems (not limited to hierarchical trees of terms). They may be creators of controlled vocabularies, thesauri, metadata schemes, or website categorization systems. “Taxonomist” is a more practical and catchy job title than “controlled vocabulary editor,” “thesaurus creator,” or “nomenclature manager.”

Yet for the scope of this book, taxonomists are not limited to people who have the word taxonomy or taxonomist within their job title. There are other job titles for essentially the same tasks, such as vocabulary developer, technical categorization analyst, and information classification specialist. There are many people who work on taxonomies as only one of several job responsibilities, whether as corporate librarians, information architects, or knowledge managers. Finally, there are those who serve in the role of taxonomist temporarily on a project, returning to other duties after completing the taxonomy.

In sum, a taxonomist is someone who creates or edits taxonomies, either singly or as part of a team of taxonomists, and taxonomies are defined as any knowledge organization system (controlled vocabulary, synonym ring, thesaurus, hierarchical term tree, or ontology) used to support information/content findability, discovery, and access. This taxonomy work may be an ongoing job responsibility or a temporary project, and it may be a primary job responsibility or a secondary responsibility. These people and those who are interested in getting into such work are the primary audience of this book.

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).<sup>2</sup> Therefore, even people with an education in information science are probably not thinking of working as a taxonomist. For this reason, too, we can say that many taxonomists become so by chance or by “accident.”

Unlike working as a reference librarian or corporate librarian, working as a taxonomist does not usually require a degree in library and information science (although it is often preferred). For this reason, too, people with varied educational backgrounds may accidentally find themselves working as taxonomists. In fact, according to the results of an online survey of taxonomists in May 2015, just about half had a master of library science (MLS) or master of library and information science (MLIS) degree. (The full survey questions and answers are reproduced in Appendix A of this book.)

Information taxonomies are relatively new and growing in terms of their applications. New web interactive technologies make taxonomies more usable and user friendly, and the exponential growth of electronic data increasingly calls for new means of organizing and accessing information. Since information taxonomies have been getting attention only since the late 1990s or around 2000, any experienced professional who is getting into taxonomies is doing so somewhat accidentally. As for entry-level taxonomy positions for the new MLIS or MIS graduate, I have yet to see such a position posted.

As for my story, although I came to developing corporate taxonomies via work on controlled vocabularies for periodical database indexing, I did come to the field of controlled vocabularies quite accidentally. I had started my career in writing and editing and then responded to a job notice for an abstractor at the computer

magazine publisher Ziff Communications, not realizing that Ziff, at the time, owned a large periodical-indexing division called Information Access Company. It turned out that the abstractors did the indexing and other metadata application as well, so after an intensive employee training on indexing, I got my first exposure to controlled vocabularies.

After indexing for a couple of years, I decided to move onward and upward into the controlled vocabulary management group and soon forgot about abstracting. But I never completely gave up writing, as the production of this book will attest. When my position was eliminated in early 2004 and I had to look for new work, I had difficulty finding a job in a profession that I didn't know what to call. My previous title had been "controlled vocabulary editor," but, alas, I found nothing by that name on the job board sites. Although publishers of aggregate periodical indexes are few and far between, it turned out that similar skills were in demand by large companies to organize and retrieve their internal documents. I then discovered *taxonomy* and *taxonomists* and realized that I could call what I had been doing for the previous 10 years "taxonomy work." With my prior taxonomist experience, I soon landed new taxonomy contract work and then, with that additional experience, a series of full-time taxonomist positions in addition to periods of independent consulting.

While taxonomy may no longer be the latest, hottest topic, as it was around 2000, it has moved beyond being a buzzword to become a topic of more stable interest. The following illustrate the sustained interest in taxonomies:

- An online discussion group dedicated to taxonomies, Taxonomy Community of Practice ([groups.yahoo.com/group/TaxoCoP](http://groups.yahoo.com/group/TaxoCoP)), has grown to 1,250 members, while a corresponding LinkedIn group has grown to 3,330 members.

## xxviii The Accidental Taxonomist

- A two-day annual conference dedicated to taxonomies, Taxonomy Boot Camp ([www.taxonomybootcamp.com](http://www.taxonomybootcamp.com)), has been growing since its inception in 2005. Registrations increased 28 percent from 229 in 2014 to 292 in 2015.<sup>3</sup> A show of hands at the opening session of the 2015 conference indicated that close to 90 percent were attending Taxonomy Boot Camp for the first time.<sup>4</sup>
- The terms *taxonomy* and *taxonomies* are appearing in increasing numbers of posted job descriptions. A search on those words in job descriptions on Monster.com found increased usage, from 160 hits for the month of April 2008 to over 1,000 for the month of October 2015.

Although there are numerous articles and conference presentations on information taxonomies, books dedicated to the subject are rare. There have been several good books published on thesaurus construction in recent decades. While these might serve as useful guides for the practicing taxonomist, thesaurus construction books do not sufficiently cover other kinds of taxonomies, such as enterprise and website taxonomies, and issues of automated indexing and search. The more recent books on taxonomies, on the other hand, are focused on enterprise taxonomies or take a more project management perspective on taxonomy creation. These may be good books for the manager or executive who is considering a taxonomy project, but they lack sufficient depth to instruct the practicing taxonomist, who needs advice on how to handle various situations in working with the taxonomy terms themselves.

What was missing, in my view, was a practical book for the person actually creating and editing the terms within a taxonomy—a resource for practicing taxonomists designed to go beyond the introductory level. Introductory information on taxonomy creation abounds in articles, conference workshops, Taxonomy Boot Camp, and a few graduate school or continuing education courses. I teach such a continuing education course myself and have been asked by

prospective students about offering an intermediate or advanced course, as nothing of that kind exists. Rather than teach a second course—an ongoing commitment—I decided to write this book.

That is not to say that *The Accidental Taxonomist* is purely at an advanced level. It is still appropriate for beginning taxonomists and includes all the content of my introductory course on creating taxonomies and controlled vocabularies. The currently practicing taxonomist will also find useful information, as additional content has been included based on various presentations and articles I have written over the past two years and on some more recent research.

Because there are many different kinds of taxonomies—for human and automated indexing, for literature retrieval and website information categorization, for consumers and internal enterprises—a taxonomist's experience in creating one kind of taxonomy is not necessarily sufficient preparation for working on a different kind of taxonomy. Thus, the book also serves the purpose of cross-training existing taxonomists for different kinds of taxonomy projects. If we want to carry the label of *taxonomist* and move from one job to another, then a broader understanding of the types of work and issues involved is needed.

The book aims to explain what you need to know to be a good taxonomist rather than how to create a taxonomy, step by step. Therefore the chapters are arranged in order of importance in terms of what you need to know, rather than in the project sequence for building a taxonomy. Chapters 1 and 2 provide background on taxonomies and taxonomists. Chapters 3 and 4 present the basics of term and relationship creation in accordance with the ANSI/NISO Z39.19 and ISO 25964 standards, which may serve as a review for experienced taxonomists but is fundamental for the new taxonomist. Chapter 5 provides practical information on the various taxonomy management software available. While some software tools have come and gone, others have been around for a long time and have staying power.

The following four chapters move beyond the basics to focus on particular issues for different types of taxonomies. Chapter 6 deals with creating taxonomies or thesauri used by human indexers, whereas Chapter 7 discusses the issues involved with creating taxonomies used in automated indexing, auto-categorization, or automated search. Chapter 8 examines various taxonomy structures, and Chapter 9 presents various display options.

Chapter 10 turns to broader issues of taxonomy planning and design, which often involve the taxonomist, and Chapter 11 deals with ongoing taxonomy work, such as the maintenance, merging, and translating of taxonomies. Finally, Chapter 12 returns the focus to the taxonomist: the nature of the work, what kind of work exists, and training and resources available.

As an aside, the quotations that lead off each chapter were proposed mottos for the Taxonomy Community of Practice discussion group ([groups.yahoo.com/neo/groups/taxocop/info](http://groups.yahoo.com/neo/groups/taxocop/info)), suggested by its various members in January 2009. (The quotation for Chapter 1 was the winning motto.)

I hope the book will prove not just informative but practical and useful as well. While it covers most of what you need to know to create taxonomies, it does not address every detail. For additional specific instructions, I highly recommend consulting the ANSI/NISO Z39.19 standard, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, which is available free of charge. It offers a wealth of information, although it is really too much for the newcomer to taxonomies to digest. That's where *The Accidental Taxonomist* comes in, and in addition, the book includes coverage of additional types of taxonomies and taxonomy features not addressed in the standard.

## Endnotes

1. Wendi Pohs, "Taxonomy and Classification Resources Should Sit on Development Teams," *Semantics in Practice Blog*, March 27, 2010, accessed December 20, 2015, [wendipohs.blogspot.com/2010/03/taxonomy-and-classification-resources.html](http://wendipohs.blogspot.com/2010/03/taxonomy-and-classification-resources.html)
2. Review of the course catalogs on the websites of the 58 academic programs listed in the directory of the American Library Association's "Alphabetical List of Institutions With ALA-Accredited Programs," accessed December 20, 2015, <http://www.ala.org/accreditedprograms/directory>
3. Taxonomy Boot Camp conference registration numbers provided by Information Today Inc. These numbers do not include an additional 50 or so speakers and exhibitors.
4. Conference chair Michael Crandall asked attendees at the opening session on November 2, 2015 if this was their first time attending, and he estimated that 90 percent of the attendees raised their hands.





## What Are Taxonomies?

*Taxonomies? That's classified information.*

—Jordan Cassel

The first step in discussing the role and work of the taxonomist is to clarify what a taxonomy is. Even if you already have some understanding of the concept, there are multiple meanings and various types of taxonomies that require further explanation. The descriptions provided here are not strict definitions, and the range of knowledge organization systems should be thought of as a spectrum.

### Definitions and Types of Taxonomies

The word *taxonomy* comes from the Greek *taxis*, meaning arrangement or order, and *nomos*, meaning law or science. For present-day information management, the term *taxonomy* is used both in the narrow sense, to mean a hierarchical classification or categorization system, and in the broad sense, in reference to any means of organizing concepts of knowledge. Some professionals do not even like to use the term, contending that it is too often ambiguous and frequently misused. Yet it has gained sufficient popularity, and a practical alternative term does not seem to exist. In this book, taxonomy will be used in its broader meaning and not limited to hierarchical structures.

In the broader sense, a taxonomy may also be referred to as a *knowledge organization system* or *knowledge organization structure*. This designation sometimes appears in scholarly discussions of the

## 2 The Accidental Taxonomist

field, in course titles at graduate schools of library and information science, and in reference to standards. The designation *knowledge organization system* was first used by the Networked Knowledge Organization Systems Working Group at its initial meeting at the Association for Computing Machinery Digital Libraries Conference in Pittsburgh, Pennsylvania, in 1998. Gail Hodge further expanded on it in an article in 2000 for the Digital Library Federation Council on Library and Information Resources. In Hodge's words,

the term *knowledge organization systems* is intended to encompass all types of schemes for organizing information and promoting knowledge management. Knowledge organization systems include classification schemes that organize materials at a general level (such as books on a shelf), subject headings that provide more detailed access, and authority files that control variant versions of key information (such as geographic names and personal names). They also include less-traditional schemes, such as semantic networks and ontologies.<sup>1</sup>

Although she does not mention taxonomies per se in this paragraph, Hodge goes on to list the various types of knowledge organization systems, which include the following:<sup>2</sup>

1. Term lists (authority files, glossaries, dictionaries, and gazetteers)
2. Classifications and categories (subject headings, classification schemes, taxonomies, and categorization schemes)
3. Relationship lists (thesauri, semantic networks, and ontologies)

The designation *knowledge organization system* has not caught on in the business world, perhaps due partially to its length. We are therefore not likely to hear of a *knowledge organization system*

*manager/editor*, but rather just a taxonomist. The designation *knowledge organization system* may be receiving wider recognition, though, with increasing adoption of the World Wide Web Consortium's recommended framework for representing these various controlled vocabularies, which is called Simple Knowledge Organization System (SKOS).

While this book uses the term *taxonomy* broadly (as a synonym for knowledge organization system), most of our discussion focuses on taxonomies that have at least some form of structure or relationship among the terms (types 2 and 3 in Hodge's list) rather than mere term lists. Indeed, people do not usually call a simple term list a taxonomy. Let us turn now to definitions and explanations of some of these different kinds of knowledge organization systems or taxonomies.

### **Controlled Vocabularies**

The term *controlled vocabulary* may cover any kind of knowledge organization system, with the possible exclusion of highly structured semantic networks or ontologies. At a minimum, a controlled vocabulary is simply a restricted list of words or terms for some specialized purpose, usually for indexing, labeling, or categorizing. It is "controlled" because only terms from the list may be used for the subject area covered. If used by more than one person, it is also controlled in the sense that there is control over who may add terms to the list and when and how they may do it. The list may grow, but only under defined policies.

The objective of a controlled vocabulary is to ensure consistency in the application of index terms, tags, or labels to avoid ambiguity and the overlooking of information if the "wrong" search term is used. When implemented in search or browse systems, the controlled vocabulary can help guide the user to where the desired information is. While controlled vocabularies are most often used in indexing or tagging, they are also used in technical writing to ensure the use of consistent language. This latter task of writing

#### 4 The Accidental Taxonomist

or creating content is not, however, part of *organizing* information. Because controlled vocabulary has this broader usage when applied to content creation, not merely information organization, the term *controlled vocabulary* should not be used as a synonym for knowledge organization system.

Most controlled vocabularies feature a *See or Use* type of cross-reference system, directing the user from one or more “non-preferred” terms to the designated “preferred” term. Only if a controlled vocabulary is very small and easily browsed, as on a single page, might such cross-referencing be unnecessary.

In certain controlled vocabularies, there could be a set of synonyms for each concept, with none of them designated as the preferred term (akin to having equivalent double posts in a back-of-the-book index instead of *See* references). This type of arrangement is known as a *synonym ring* or a *synset* because all synonyms are equal and can be expressed in a circular ring of interrelationships. An example of a synonym ring, as illustrated in Figure 1.1, is the series of terms *applications*, *software*, *computer programs*, and *tools*. Synonym rings may be used when the browsable list of terms or entries is not displayed to the user and when the user merely accesses the terms via a search box. If the synonyms are used behind the scenes with a search engine and never displayed as a browsable list for the user, the distinction between preferred and nonpreferred terms is thus moot. Though these types of controlled vocabularies are quite common, they are often invisible to the user, so the terminology (synonym ring and synset) is not widely known.

Sometimes controlled vocabularies are referred to as *authority files*, especially if they contain just named entities. Named entities are proper-noun terms, such as specific person names, place names, company names, organization names, product names, and names of published works. These also require control for consistent formats, use of abbreviations, spelling, and so forth.

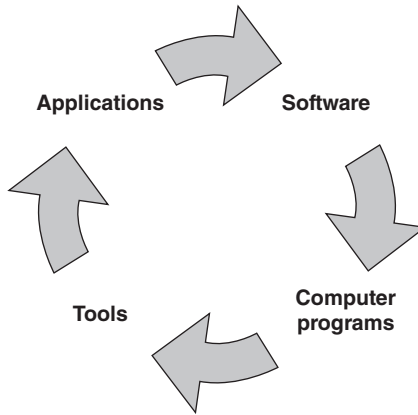


Figure 1.1 Example of terms in a synonym ring

Controlled vocabularies may or may not have relationships among their terms. Simple controlled vocabularies, such as a temporary offline list created by an indexer to ensure consistent indexing or a synonym ring used behind the scenes in a search, do not have any structured relationships other than preferred and non-preferred terms. Other controlled vocabularies may have broader/narrower and related term relationships and still be called controlled vocabularies rather than thesauri or taxonomies. This is often the case at periodical and reference index publishers, such as Gale (a part of Cengage Learning), EBSCO, and ProQuest, which maintain controlled vocabularies for use in their periodical indexes. In some cases, the publisher maintains multiple kinds of controlled vocabularies, some being more structured than others, and controlled vocabulary is the more generic designation for all of these.

### **Hierarchical Taxonomies**

When we think of taxonomy, hierarchical classification systems are what typically come to mind. However, as explained in the previous section, we are using a broader definition of taxonomy that encompasses all kinds of knowledge organization systems. So taxonomies

## 6 The Accidental Taxonomist

that are structured as hierarchies will be referred to specifically as *hierarchical taxonomies*.

A hierarchical taxonomy is a kind of controlled vocabulary in which each term is connected to a designated broader term (unless it is the top-level term) and one or more narrower terms (unless it is the bottom-level term), and all the terms are organized into a single large hierarchical structure. Taxonomy in this case could apply to a single hierarchy or a limited set of hierarchies. This type of structure is often referred to as a *tree*, with a trunk, main branches, and more and more smaller branches off the main branches. Actually, if the taxonomy is displayed as a tree, it is an upside-down tree, with multiple smaller branches for narrower terms lower down on the page or screen. Another way to describe such structure is a taxonomy with *nested categories*. The expression *to drill down* is often used to describe how a user navigates down through the branches. An example of an excerpt from a hierarchical taxonomy appears in Figure 1.2.

The classic example of a hierarchical taxonomy is the Linnaean taxonomy (named after Carolus Linnaeus) of biological organisms, with the hierarchical top-down structure: kingdom, phylum, class, order, family, genus, and species. Hierarchical taxonomies are also common in geospatial classification, as for regions, countries, provinces, and cities. While hierarchical taxonomies tend to be used mostly for generic things or concepts, they can also be used for proper nouns that naturally fall into a hierarchy, such as place names, product names, government agency names, or corporate department names.

The structure of a hierarchical taxonomy often reflects an organization of nested categories. Some hierarchical taxonomies permit a term to have multiple broader terms, thus appearing in multiple places in the taxonomy, whereas other hierarchical taxonomies do not permit this “polyhierarchy” structure. Hierarchical taxonomies may or may not make use of nonpreferred terms. Finally,

<b>Top Level Headings</b>	
Business and industry	Leisure and culture
Economics and finance	. Arts and entertainment venues
Education and skills	. . Museums and galleries
Employment, jobs and careers	. Children's activities
Environment	. Culture and creativity
Government, politics and public administration	. . Architecture
Health, well-being and care	. . Crafts
Housing	. . Heritage
Information and communication	. . Literature
International affairs and defence	. . Music
Leisure and culture	. . Performing arts
Life in the community	. . Visual arts
People and organisations	. Entertainment and events
Public order, justice and rights	. Gambling and lotteries
Science, technology and innovation	. Hobbies and interests
Transport and infrastructure	. Parks and gardens
Leisure and culture	. Sports and recreation
	. . Team sports
	. . . Cricket
	. . . Football
	. . . Rugby
	. . Water sports
	. . Winter sports
	. Sports and recreation facilities
	. Tourism
	. . Passports and visas
	. Young people's activities

Figure 1.2 Terms in an expandable hierarchical taxonomy; top categories (left) and the expansion of one category (right), from the Abridged Integrated Public Sector Vocabulary, Version 2.00 ([www.esd.org.uk/standards/ipsv](http://www.esd.org.uk/standards/ipsv))

nonhierarchical related term relationships may exist but usually are not present in such hierarchical taxonomies.

In contrast to the other types of taxonomies described subsequently in this chapter and this book, the hierarchical taxonomy is actually not a defined type of taxonomy. Rather, it is my designation for the narrower, standard definition of taxonomy: “A collection of controlled vocabulary terms organized into a hierarchical structure.”<sup>3</sup> It is a kind of taxonomy that is commonly seen in countless real-world applications. And it is the type of taxonomy that the accidental taxonomist is probably most likely to create.

While in some contexts a faceted taxonomy may be considered different from a hierarchical taxonomy, faceted taxonomies, at the



## **8 The Accidental Taxonomist**

high level, are variants of hierarchical taxonomies that are implemented and used in a special way. Faceted taxonomies, like other hierarchical taxonomies, are intended to be browsed by the end user, starting from the top down. A facet is like a hierarchy, and the facet name is like the top-level term in a hierarchy. Relationships between terms that are other than hierarchical are optional and less common. Facets are explained in the section on Retrieval Support later in this chapter and are discussed in much more detail in Chapter 8.

### ***Alpha-Numeric Classification Systems***

Classification systems that utilize numeric, alphabetical, or alpha-numeric codes along with the descriptive terms are a type of hierarchical taxonomy. Examples include the Dewey Decimal Classification system and the Library of Congress Classification System for cataloging books. Other well-known examples of hierarchical taxonomies are the Standard Industrial Classification (SIC) and North American Industrial Classification Systems (NAICS) codes for classifying industries. There are classification systems from the United Nations and international agencies, such as the United Nations Statistics Division's classification for activities, products, and expenditure purposes, and the International Press Telecommunications Council's (IPTC) NewsCodes.

Classification systems, however, are not used exactly the same way as purely topical taxonomies. The conceptual idea is to “classify” documents, or put them into “classes,” preferably in only one place. (In the case of classification of physical objects—for the purpose of shelving books, for example—there can be only one place.) Classes are like metadata for the document: They tell what class the document belongs in. They are not for topics discussed in the documents or what the document is all about; that is the purpose of indexing. A document could receive topical indexing from a taxonomy that is supplemental to its classification.

Furthermore, numeric code-based systems are not flexible and cannot easily be changed. It is not usually practical to insert additional codes into the scheme, unless perhaps the system allows for one additional hierarchical level. Because these systems are relatively unchanging, they don't need to be created or updated often, and their applications are somewhat limited. Thus, most of the subject areas that could use classification systems already have them, and they don't need changing. And those subject areas that don't have them are not suitable for them. What this means is that there is not much work for taxonomists in the area of classification systems.

### **Thesauri**

The classic meaning of a *thesaurus* is *a kind of dictionary*, such as *Roget's*, that contains synonyms or alternate expressions (and possibly even antonyms) for each term entry. A thesaurus for information management and retrieval shares this characteristic of listing similar terms at each controlled vocabulary term entry. The difference is that a dictionary-thesaurus includes all the associated terms that *could potentially* be used in place of the term entry in various contexts; the user (often a writer) needs to consider the specific context in each case because in certain contexts some of the alternate terms would not be appropriate. The information retrieval thesaurus, on the other hand, is designed for use in *all* contexts within the domain of content covered, regardless of any specific term usage or document. The synonyms or near synonyms must therefore be suitably equivalent in *all* circumstances. An information retrieval thesaurus must clearly specify which terms can be used as synonyms (used from), which are more specific (narrower terms), which are broader terms, and which are merely related terms.

A thesaurus, therefore, is a more structured type of controlled vocabulary that provides information about each term and its

## 10 The Accidental Taxonomist

relationships to other terms within the same thesaurus. The leading standards that provide guidance for creating such thesauri are the following:

- International Organization for Standardization ([www.iso.org/iso/catalogue\\_ics](http://www.iso.org/iso/catalogue_ics)) ISO 25964. *Information and documentation - Thesauri and interoperability with other vocabularies*  
Part 1 (2011): *Thesauri for information retrieval*  
[www.iso.org/obp/ui/#iso:std:iso:25964:-1:ed-1:v1:en](http://www.iso.org/obp/ui/#iso:std:iso:25964:-1:ed-1:v1:en)  
Part 2 (2013): *Interoperability with other vocabularies*  
[www.iso.org/obp/ui/#iso:std:iso:25964:-2:ed-1:v1:en](http://www.iso.org/obp/ui/#iso:std:iso:25964:-2:ed-1:v1:en)

These have also been adopted by the British Standards Institute as BS ISO 25964-1:2011 and BS ISO 25964-2:2013.

- American National Standards Institute and National Information Standards Organization ([www.niso.org](http://www.niso.org)); ANSI/NISO Z39.19 2005 (R2010): *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*  
[www.niso.org/standards/z39-19-2005](http://www.niso.org/standards/z39-19-2005)

Although the ANSI/NISO standard refers to “controlled vocabularies,” a vocabulary created in accordance with these guidelines is usually called a thesaurus.

The standards explain in detail the three types of relationships in a thesaurus: hierarchical (broader term/narrower term), associative (related term), and equivalence (use/used for). Additional information about a term, such as a *scope note*, may be included to clarify usage. An example of a term and its details from a thesaurus is shown in Figure 1.3. The consensus is that if a controlled vocabulary includes both broader/narrower and related term relationships between terms, along with non-preferred terms that redirect to the accepted term, then it is called a thesaurus.

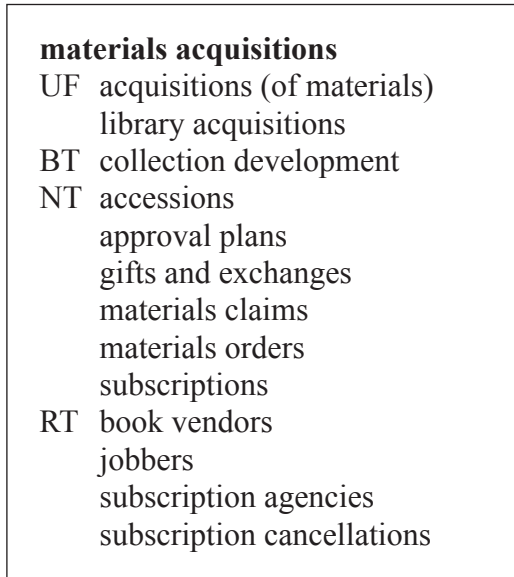


Figure 1.3 A term in the *ASIS&T Thesaurus* with its various relationships to other terms (BT: broader term; NT: narrower term; RT: related term; UF: used from)<sup>4</sup>

In comparing a thesaurus with a hierarchical taxonomy, a thesaurus typically includes the features of a taxonomy plus the additional feature of associative relationships, for a greater degree of structural complexity. However, while all terms must belong to a limited number of hierarchies within a hierarchical taxonomy, this is not a strict requirement for a thesaurus. Although most thesaurus entries will list a broader and/or a narrower term, such relationships are not necessarily required for every term. If there is no appropriate broader term, that relationship may be omitted. In a thesaurus, the focus is more on the individual terms than on the top-down structure. Thus a thesaurus might include multiple small hierarchies, comprising as few as two or three terms, without the strong overarching tree structure typical of a hierarchical taxonomy.

If you had to force all the terms in a thesaurus into a single hierarchical tree, some of the hierarchical relationships would probably

## 12 The Accidental Taxonomist

be imperfect. Thesaurus guidelines, however, mandate that each term's hierarchical relationships be accurate and valid. In addition, having multiple broader terms for an entry is never a problem in a thesaurus, whereas such "polyhierarchies" may be prohibited in a given hierarchical taxonomy. Some thesauri do in fact have a significant hierarchical structure, and thus the distinction between a hierarchical taxonomy and a thesaurus may be blurred. Finally, recursive retrieval by a broader term (explained in Chapter 9) is not as common in a thesaurus as in a hierarchical taxonomy.

The greater detail and information contained in a thesaurus, compared with a simple controlled vocabulary or a hierarchical taxonomy, aids the user (whether the indexer or the searcher) in finding the most appropriate term more easily. A thesaurus structure is especially useful for a relatively large controlled vocabulary that involves human indexing and/or supports a term list display that the end user (searcher) can browse. In contrast to a hierarchical taxonomy, which is designed for user navigation from the top down, a thesaurus with multiple means of access can more easily contain a greater number of terms. Thus, a thesaurus may be able to support more granular (specific) and extensive indexing than a simple hierarchical taxonomy can, especially if the hierarchical taxonomy lacks nonpreferred terms. As thesauri explain relationships among terms, they are more common in specialized subject areas, where the purpose is not merely to aid the user in finding information but also to aid the user in obtaining a better understanding of the terminology. In some cases, thesauri have even been published and printed as stand-alone works, separate from any indexed content.

Examples of thesauri include the Getty Art & Architecture Thesaurus ([getty.edu/research/tools/vocabularies/aat](http://getty.edu/research/tools/vocabularies/aat)), the ERIC (Education Resources Information Center) Thesaurus for education research ([eric.ed.gov](http://eric.ed.gov)), and the NASA Thesaurus of aeronautics and space terminology ([www.sti.nasa.gov/thesvol1.pdf](http://www.sti.nasa.gov/thesvol1.pdf)). The

periodical and reference index publisher ProQuest also refers to its topical controlled vocabulary as a thesaurus.

### **Ontologies**

An ontology can be considered a type of taxonomy with even more complex relationships between terms than in a thesaurus. Actually, an ontology is more than that; it aims to describe a domain of knowledge, a subject area, by both its terms (called *individuals* or *instances*) and their relationships and thus supports inferencing. This objective of a more complex and complete representation of knowledge stems from the etymology of the word *ontology*, which originally meant the study of the nature of being or existence. Tom Gruber provides a current definition of ontology:

An ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. ... ontology can be viewed as a level of abstraction of data models, analogous to hierarchical and relational models.<sup>5</sup>

The relationships between terms within an ontology are not limited to broader/narrower and related. Rather, there can be any number of domain-specific types of relationship pairs, such as owns/belongs to, produces/is produced by, and has members/is a member of. The creator of the ontology also creates these relationship types. Thus, not only do the terms have meanings, but also the relationships themselves have meanings. Relationships with meanings are called semantic relationships.

The terms within an ontology not only have simple descriptions, such as scope notes in a thesaurus, but are also accompanied by specific attributes in a more structured format, such as properties, features, characteristics, or parameters. The terms also have assigned classes, which the ontologist defines, as an additional kind of classification. All of these components of an ontology—semantic

## 14 The Accidental Taxonomist

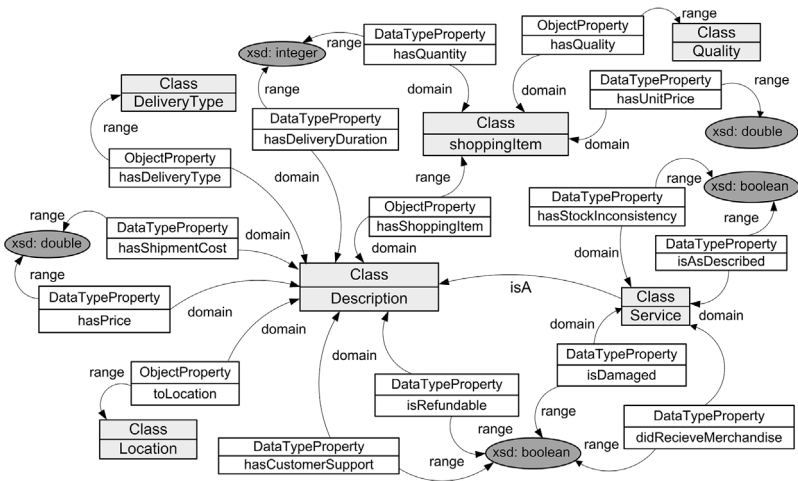


Figure 1.4 Example of a domain ontology dealing with retail management (reproduced with permission of the creators, Murat Sensoy and Pinar Yolum)

relationships, attributes (for each of the terms/instances), and classes—contribute to making an ontology a richer source of information than a mere hierarchical taxonomy or thesaurus. A schematic representation of part of an ontology dealing with retail management appears in Figure 1.4.

While not considered standards, there are guidelines of specifications for constructing ontologies in machine-readable format for the web, which has become the most common implementation of this type of taxonomy. The World Wide Web Consortium (W3C) has published the RDF (resource description framework) Schema and the Web Ontology Language (OWL) recommendation. There is also a presentation format for ontologies called Topic Maps, which is the set of ISO 13250 standards (six parts). Topic Maps are implemented more in Europe than in North America. A looser structure of knowledge organization that does not attempt to adhere to such guidelines might be called a semantic network instead of an actual ontology.

Ontologies are suitable for any subject area, but a significant percentage of those currently published have been in the biological sciences, such as the Gene Ontology, Protein Ontology, Systems Biology Ontology, and Purdue Ontology for Pharmaceutical Engineering. It is an interesting irony that taxonomies, which got their start in biological classification, are now widely used for any form of knowledge, while ontologies, which originally applied to the broad scope of existence, are now used most often in the field of biology. As other scientists find a need to express more complex relationships among terms in their disciplines, the spread of ontologies to other subject areas, however, will likely increase. There is also a growing importance of ontologies in semantic search engine deployment in specialized industries, and building ontologies could be a growth area for experienced taxonomists. In 2009, a new organization for supporting ontologies, the International Association for Ontology and Its Applications ([www.iaoa.org](http://www.iaoa.org)), was founded.

The designation given to a knowledge organization system—controlled vocabulary, taxonomy, thesaurus, ontology, and so on—depends largely on the complexity of the structure, but complexity is not the only factor to be considered. As all these designations have ambiguous meanings, the choice of what to call a set of terms also depends on what is most clear and understandable to the contributors, stakeholders, or end users. Depending on the display of the knowledge organization system, the end users may not even need to know what it is called. The confusion in terminology is why we default to using the single designation of taxonomy in most contexts.

## **Taxonomies and Metadata**

There is significant overlap between taxonomy and metadata. Metadata, sometimes called “data about data,” is all the recorded, structured information about a content item, such as a document, digital asset (such as a video or image file), or webpage. Taxonomies



## **16 The Accidental Taxonomist**

(or more generally, controlled vocabularies) are often, but not always, metadata. And much, but not all, of metadata utilizes controlled vocabularies.

A content repository has a metadata schema, which may follow a suggested standard, such as Dublin Core Metadata Elements or MARC (Machine-Readable Cataloging), which is useful if the content is shared among different organizations; or it may have a customized metadata schema, which is more practical for managing metadata within an organization. A metadata schema comprises both the defined set of metadata elements or fields and the rules for each of those fields. Fields get filled or “populated” with specific values as appropriate for each individual content item.

Different types of metadata serve different purposes. The National Information Standards Organization (NISO) defines three kinds of metadata: descriptive, structural, and administrative. Descriptive metadata includes information on what a resource is about, expressed in keywords or short descriptions; and also includes other descriptive information that could be used to look up and retrieve the item, such as title, author, and document type. Structural metadata describes features of the resource, such as pagination or size. Administrative metadata describes information needed to manage a resource, such as its creation date, access rights, intellectual property rights, and archival preservation information. There are other methods besides NISO for classifying metadata types, but most methods distinguish between metadata for managing content and metadata for aiding in search or discovery and retrieval of content.

### **Taxonomy as Part of Metadata**

Taxonomies, or controlled vocabularies in general, are associated with the descriptive type of metadata, for two reasons. First, taxonomists, by the nature of their work, are focused on the goal of descriptive metadata, which is to help users find content. Second, descriptive metadata tends to use controlled vocabularies

more than other types of metadata. While administrative or structural metadata may require controlled lists of terms to populate some of its fields, a controlled “list” is not necessarily a controlled “vocabulary.”

Regardless of the type of metadata, (descriptive, structural, or administrative), a specific metadata field may either allow free text or require the user to select from a controlled list of options. A controlled vocabulary is, of course, a type of controlled list. It would be incorrect, however, to call all controlled lists “controlled vocabularies.” For example, the controlled list for a metadata field may consist of just a pair of values, such as *yes* or *no*, *male* or *female*, or *new* or *used*, or it may consist of just three or four values, such as *small*, *medium*, and *large*. These types of lists are not controlled vocabularies, because part of the definition of a controlled vocabulary is that a term is designated for a concept. In the preceding examples, there is not really any question of what word or term to use for each concept in the list of two or three. We don’t even use the word “term” but rather “value” to refer to the choices in short controlled lists. There is no rule, however, as to the number of values a list must have before it is considered a controlled vocabulary. Rather, the defining feature of a controlled vocabulary is that concept-naming decisions need to be made when developing the list. Therefore, the list of 50 states of the United States by itself is not a controlled vocabulary, just a list, because there is no question of what to call them, whereas a list of just 15 “activity types” would likely be a controlled vocabulary, because names for those activities require research and consideration before a choice can be made from among various alternatives.

Controlled vocabularies of any size, including hierarchical taxonomies, may be used to support one or more descriptive metadata fields, especially a field that is called Subject, Topic, or Descriptor. A taxonomist is not necessarily responsible for all metadata, so he or she needs to work in collaboration with a metadata architect,

## **18 The Accidental Taxonomist**

metadata librarian, or content architect, especially in the blurred area of responsibility between short controlled vocabularies and long controlled lists. In addition to determining the metadata fields and their values, other decisions need to be made: whether assigning/tagging values from a specific metadata field is required or optional, whether a metadata field may hold only one value or can permit multiple values, and whether the field will be displayed in the user interface for end-user search-and-retrieval purposes.

### **Taxonomy That Is Not Metadata**

If a taxonomy is implemented in a way that the terms, unlike other metadata, are *not* attached to a content item, then the taxonomy might not be part of metadata. If a taxonomy is implemented to support dynamic auto-indexing or search, and executed “on the fly,” rather than being permanently attached to a record, then it is not metadata. While we have already stated that a hierarchical navigation design is not necessarily a taxonomy, in some cases a navigation scheme could be a taxonomy and may not be metadata. If a navigation scheme were designed as a taxonomy and taxonomy terms were hyperlinked to content, with some terms indirectly linked to more than one content item (such as through an intermediate navigation page), then the taxonomy terms would not necessarily be metadata.

### **Applications and Purposes of Taxonomies**

As we have seen from the various definitions, there are different kinds of taxonomies or controlled vocabularies, based on their complexity. However, complexity is only one way to classify taxonomies. A more practical approach is to categorize them by their application and use. While a particular taxonomy can certainly serve multiple functions, there tends to be a certain emphasis in its design, use, and purposes. As such, each taxonomy serves primarily one of the following three functions, although there certainly can be combinations of the different types:

1. Indexing support
2. Retrieval support
3. Organization and navigation support

### **Indexing Support**

For indexing or cataloging support, a taxonomy, better known as a controlled vocabulary in this context, is a list of agreed-on terms for the human indexing or cataloging of multiple documents and/or for indexing performed by multiple indexers, to ensure consistency. If multiple documents, especially by different authors, will be indexed over time, the indexer is apt to forget exactly which index terms were assigned and perhaps inadvertently use different synonyms when the same topic comes up in a different document. Similarly, different indexers will also choose different index terms for the same topic if not forced to use a controlled vocabulary.

Thus, the taxonomy's initial purpose is to serve the people doing the indexing, although a second, equally important purpose is to serve the end users, who, of course, benefit from consistently indexed content and may also have access to the taxonomy. This type of controlled vocabulary is used for cataloging entire works and for indexes to periodical articles, image files, database records, multivolume printed works, webpages, and so on. Because indexers must always choose the most accurate terms, they often use a more structured thesaurus type of controlled vocabulary. The broader, narrower, and related term relationships guide the indexer to the best term, and scope notes further clarify ambiguous terms. Named entities are often indexed, too, and these are managed in an authority file. An authority file lacks the interterm relationships of a thesaurus but may have many synonymous nonpreferred terms for each preferred term, such as variations on an individual's name.

Controlled vocabularies for indexing support have been around the longest, and their format may be electronic or print. Such controlled vocabularies are used by reference and periodical article

## 20 The Accidental Taxonomist

database publishers, including Gale (a part of Cengage Learning), EBSCO, and ProQuest, in more specialized subject databases, such as Chemical Abstracts and PsycINFO. And they are also used in the internal documents of large companies, especially those in the sciences. The fact that some of these controlled vocabularies are offered for sale/license illustrates the fact that they serve the purpose of indexing and not just specific content retrieval.

While controlled vocabularies for indexing are quite widespread, those that are publicly available on the web are limited and tend to be those published by public agencies. You may search or browse them, and in some cases, you may also access linked content. Library of Congress Subject Headings and Medical Subject Headings are two such examples.

Library of Congress Subject Headings (LCSH) ([authorities.loc.gov](http://authorities.loc.gov)) contains both subjects and names and covers all subject areas. LCSH was originally established for cataloging library materials but has also been adopted by various publishers for indexing articles. The terms are called *authorities*, as in authority file, even those that are not named entities. The purpose of the website is to aid catalogers of library materials in finding the approved subject heading in the Library of Congress controlled vocabulary. It is not aimed at the end user looking for a book, although consistently cataloged books will, of course, benefit the user. The subject headings can be searched and the results browsed alphabetically. Nonpreferred terms are included in the alphabetical list along with preferred terms. Nonpreferred terms are prefaced by a button labeled References, which provides a cross-reference to the preferred term. Preferred terms are called *authorized headings* (see Figure 1.5).

Medical Subject Headings (MeSH) ([www.nlm.nih.gov/mesh/MBrowser.html](http://www.nlm.nih.gov/mesh/MBrowser.html)) is the thesaurus of the US National Library of Medicine, which is considered the authority for medical terms. Users can search terms, or they can browse by selecting the button

The Library of Congress >> Go to Library of Congress Online Catalog

---

**LIBRARY OF CONGRESS AUTHORITIES**

[Help](#) | [New Search](#) | [Search History](#) | [Headings List](#) | [Start Over](#)

SOURCE OF HEADINGS: Library of Congress Online Catalog  
 YOU SEARCHED: Subject Authority Headings = world wide web  
 SEARCH RESULTS: Displaying 1 through 100 of 100.

◀ Previous    Next ▶


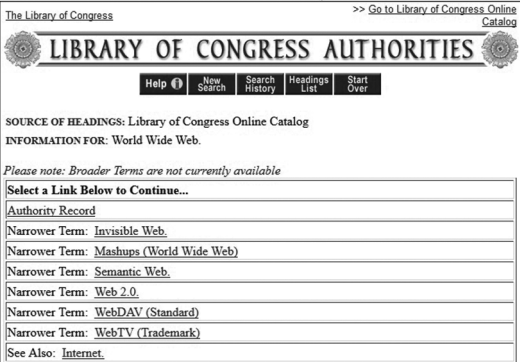
#	Bib Records	<i>select icon in first column to...</i> View Authority Headings/References	Type of Heading
 1	1253	World Wide Web.	Library of Congress subject headings
2	29	World Wide Web.	Schlagwortnormdatei
3	17	World Wide Web.	LC Annotated Children's Catalog subject headings
4	13	World Wide Web.	
5	5	World wide web.	
6	4	World Wide Web.	
7	3	World Wide Web.	
8	2	World Wide Web.	
9	1	WORLD WIDE WEB	
10	1	WORLD WIDE WEB.	
11	13	World Wide Web 2.0.	
12	5	World Wide Web 2.0.	
13	1	World Wide Web 2.0--RS	

Figure 1.5 Two successive screenshots from Library of Congress Subject Headings, searching on the term *World Wide Web* and displaying its details

Navigate from Tree Top. The browse display is hierarchical rather than alphabetical. Clicking once on a term expands the tree and reveals its narrower terms; double-clicking on a term displays its details (see Figure 1.6).

Other examples of thesauri that aid indexing and are publicly available include the ERIC Thesaurus (eric.ed.gov), sponsored by the Institute of Education Sciences of the US Department of Education, and the various controlled vocabularies of the Getty Research Institute of the J. Paul Getty Trust: the Getty Art & Architecture Thesaurus, Getty Thesaurus of Geographic Names, Cultural Objects Name Authority, and Union List of Artist Names (getty.edu/research/tools/vocabularies).

## 22 The Accidental Taxonomist

<b>MeSH Heading</b>	Arm Injuries
<b>Tree Number</b>	C26.088
<b>Annotation</b>	GEN or unspecified; consider also / <a href="#">inj</a> with specific bones of arm; also available are <a href="#">FOREARM INJURIES</a> ; <a href="#">HAND INJURIES</a> ; <a href="#">FINGER INJURIES</a> ; <a href="#">WRIST INJURIES</a> & many specific organ/fract precoords
<b>Scope Note</b>	General or unspecified injuries involving the arm.
<b>Entry Term</b>	Injuries, Arm
<b>Allowable Qualifiers</b>	BL CF CI CL CN CO DH DI DT EC EH EM EN EP ET GE HI IM ME MI MO NU PA PC PP PS PX RA RH RI RT SU TH UR US VE VI
<b>Entry Version</b>	ARM INJ
<b>Date of Entry</b>	19990101
<b>Unique ID</b>	D001134

<b>MeSH Tree Structures</b>	
<a href="#">Wounds and Injuries [C26]</a>	
<a href="#">Abdominal Injuries [C26.017] +</a>	
<a href="#">Amputation, Traumatic [C26.062]</a>	
▶ <a href="#">Arm Injuries [C26.088]</a>	
	<a href="#">Forearm Injuries [C26.088.2681] +</a>
	<a href="#">Humeral Fractures [C26.088.390]</a>
	<a href="#">Shoulder Dislocation [C26.088.666]</a>
	<a href="#">Shoulder Fractures [C26.088.749]</a>
	<a href="#">Tennis Elbow [C26.088.890]</a>
	<a href="#">Wrist Injuries [C26.088.906]</a>
<a href="#">Asphyxia [C26.103]</a>	
<a href="#">Athletic Injuries [C26.115]</a>	
<a href="#">Back Injuries [C26.117] +</a>	
<a href="#">Burns [C26.120] +</a>	

Figure 1.6 Two screenshots from Medical Subject Headings, showing the term record and location in the tree structure for the selected term *arm injuries*

### Retrieval Support

A taxonomy that serves indexing also serves end-user retrieval. Searchers benefit from nonpreferred terms, as their search terms may be different from the terms used to index the document. For example, a user may type in **doctors** for articles that are about physicians. Users can also take advantage of broader and narrower term relationships or hierarchies to broaden or narrow their search. These relationships, and also the related term relationships, may suggest to users other possible terms of interest. In such cases, the end-user searcher will be seeing an explicit representation of the taxonomy for navigation.

There are also taxonomies designed to aid search retrieval without supporting human indexing. These taxonomies are typically mapping tables of terms and their synonyms/variants designed to aid online retrieval. These might be synonym rings or synsets, especially if the terms are not even displayed to the user, or if there is a display, it may designate preferred terms.

Depending on the user interface display, there may or may not be a hierarchical structure to the taxonomy. A hierarchical arrangement allows users to browse and locate narrower (more specific) subjects of interest. Thus, users find out what is included in the taxonomy and what is not, saving themselves the trouble of repeatedly typing in terms that yield no results. Users may also find related subjects of interest by browsing the hierarchies.

These types of controlled vocabularies are often used with website search engines, enterprise search systems (used internally within a large organization), online databases, and large commercial directories (such as online “yellow pages” or classified ads). The format is always electronic, and a form of automated indexing is usually involved.

Examples of taxonomies aiding retrieval include Verizon SuperPages’s yellow pages directory site ([www.superpages.com/yellowpages](http://www.superpages.com/yellowpages)) and the Amazon.com ecommerce site’s “Shop by Department” categories ([www.amazon.com/gp/site-directory](http://www.amazon.com/gp/site-directory)), as shown in Figure 1.7.

While a hierarchy can be selected for browsing in each, the synonyms in the case of Verizon SuperPages and the related subject links in the case of Amazon.com are not displayed to the user, although the links are evident in the display of results.

### **Faceted Taxonomies for Retrieval Support**

One way to better serve specifically the retrieval of data is to construct a controlled vocabulary that is divided into multiple subsets, lists of terms of different types representing different aspects of information. These aspects are often called *facets*, and this type of



## 24 The Accidental Taxonomist

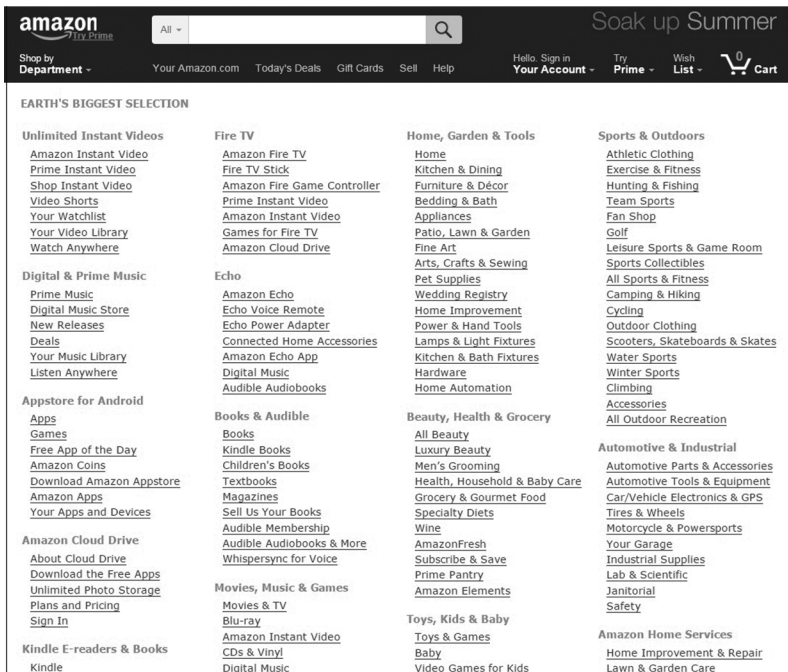


Figure 1.7 Top-level taxonomy of Amazon.com

controlled vocabulary is therefore called a *faceted taxonomy*. Examples of facets might be people, places, events, products, and laws. Facets can also reflect metadata other than subject categories, such as document type, author, and audience. The search interface for a faceted taxonomy is designed for the user to search on a selected combination of multiple facets.

Faceted taxonomies are commonly used for online databases and ecommerce sites, such as the shoe-retailing site Shoebuy.com. In Shoebuy's (www.shoebuy.com) advanced search, the facets are Category, Size, Width, Brand, Color, Price Range, and, additionally for women's shoes, Heel (height). Another example of a faceted browse interface can be found on the Microbial Life Education Resources site (serc.carleton.edu/microbelife/resources), where

<p>Results 1 - 10 of 796 matches</p> <p><b>A Natural View of Microbial Biodiversity within Hot Spring</b>  <a href="#">Cyanobacterial Mat Communities</a> part of SERC Web Resource Collection  <a href="http://mmbbr.asm.org/cgi/content/full/62/4/1353">http://mmbbr.asm.org/cgi/content/full/62/4/1353</a>                  This 18-page review summarizes a decade of research studying hot spring cyanobacterial mats as models for understanding principles of microbial community ecology. It addresses the emergence of ...                  Subject: Biology: Biology:Evolution: <b>Bioinformatics, Patterns</b>, Biology:Ecology: Symbiotic Relations: <b>Mutualism</b>, Biology:Diversity, Molecular Biology, Microbiology:Methods of Microbiology: <b>Molecular Methods</b>, Biology:Microbiology, Ecology: <b>Principles, Biofilms</b>                  Resource Type: Scientific Resources: <b>Overview/Reference Work</b>                  Extreme Environments: <b>Extremely Hot</b>                  Grade Level: <b>Graduate/Professional, College Upper (15-16)</b></p> <hr/> <p><b>Light-Induced Motility of Thermophilic <i>Synechococcus</i> Isolates from Octopus Spring, Yellowstone National Park</b> part of SERC Web Resource Collection  <a href="http://aem.asm.org/cgi/content/abstract/63/16/2247">http://aem.asm.org/cgi/content/abstract/63/16/2247</a>                  This journal abstract reports that light-induced motility of different cyanobacterial lineages may be indicative that the trait is possibly widespread among the group. The full article, in PDF ...                  Subject: Biology: Biology: <b>Molecular Biology, Microbiology</b>, Microbiology:Methods of Microbiology: <b>Microscopy</b>, Biology:Ecology: Symbiotic Relations: <b>Mutualism</b>, Biology:Evolution: <b>Patterns</b>                  Resource Type: Scientific Resources: <b>Research Lists</b>                  Extreme Environments: <b>Alkaline, Extremely Hot</b>                  Grade Level: <b>Graduate/Professional, College Upper (15-16)</b></p> <hr/> <p><b>The Astrobiology Curriculum</b> part of SERC Web Resource Collection  <a href="http://astrobio.terc.edu/">http://astrobio.terc.edu/</a>                  This site is a preview version of the Astrobiology Curriculum, an interdisciplinary year-long course for middle and high school students developed by TERC and NASA. Through a series of investigations</p>	<p><b>Refine the Results</b> ↴</p> <p><b>Subject: Biology</b>                  277 matches <b>General/Other</b>                  Astrobiology 95 matches                  Biogeochemistry 125 matches                  Diversity 135 matches                  Ecology 596 matches                  Evolution 206 matches                  Microbiology 796 matches                  Molecular Biology 169 matches</p> <p><b>Resource Type</b>                  Activities 125 matches                  Assessments 12 matches                  Course Information 20 matches                  Datasets and Tools 27 matches                  Audio/Visual 153 matches                  Computer Applications 18 matches                  Pedagogic Resources 61 matches                  Scientific Resources 682 matches                  Biographical Resources 4 matches                  Policy Resources 13 matches</p> <p><b>Extreme Environments</b>                  Alkaline 52 matches                  Acidic 55 matches                  Extremely Cold 52 matches                  Extremely Hot 116 matches                  Hypersaline 61 matches                  High Pressure 59 matches                  High Radiation 24 matches                  Anhydrous 32 matches                  Anoxic 67 matches                  Altered by Humans 65 matches</p> <p><b>Ocean Environments</b>                  Coastal and Estuarine 159 matches                  Shallow Sea Floor/Continental Shelf 31 matches                  Deep Sea Floor/Abyssal 41 matches                  Surface Waters 79 matches                  Deep Waters 28 matches</p> <p><b>Grade Level</b></p>
---	---

Figure 1.8 Faceted taxonomy in the right margin of the Microbial Life Education Resources search site

facets are Subject, Resource Type, Extreme Environments, Ocean Environments, and Grade Level (Figure 1.8).

Faceted taxonomies, or *faceted browse* systems, make use of the electronic format. Depending on the size of the vocabulary in each facet, these taxonomies may or may not make use of synonyms and may or may not have hierarchies within them. Some facets can be quite small. Facets are discussed in more detail in Chapter 8.

### Organization and Navigation Support

A taxonomy, as a hierarchy, can provide a categorization or classification system for things or for information. For the organization of information, we often see taxonomies applied in website information

## 26 The Accidental Taxonomist

architecture (structural design), online information services, intranet content organization, and corporate content management systems. In such websites or enterprise taxonomies, the emphasis is on classification and guided user navigation rather than on search and retrieval of specific information. *Navigation* means finding one's way around, whereas *retrieval* means going after specific information. The taxonomy for a website is a lot like a table of contents, organized by topic. It can be reflected in the navigational menu and in the site map. As such, it might be called a *navigational taxonomy*. These types of taxonomies tend to be relatively small and can coexist with additional, more detailed taxonomies elsewhere on the website.

An example of navigational website taxonomy that is present in both the site map and the navigational menu can be found on the Information Architecture Institute site map ([iainstitute.org/en/site-map.php](http://iainstitute.org/en/site-map.php)), where the top-level categories of the taxonomy and the navigation are Member Services, IA Network, Learning IA, and About Us (Figure 1.9).

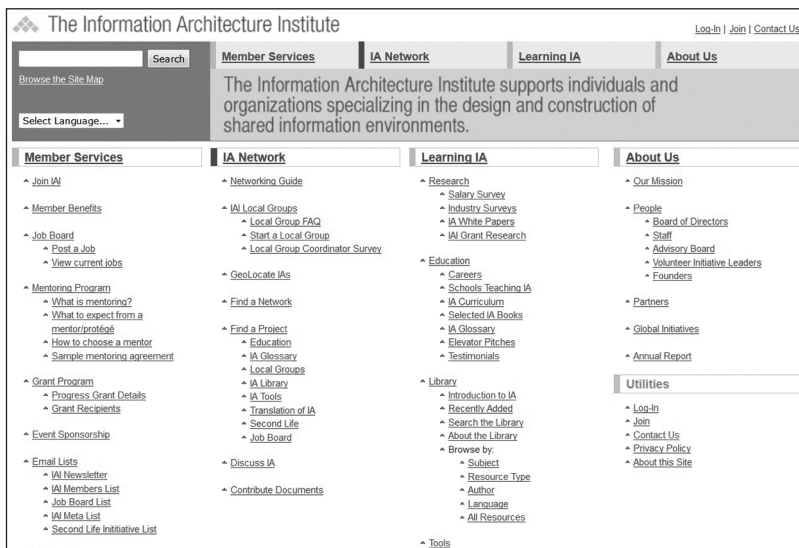


Figure 1.9 The Information Architecture Institute site map, a navigational taxonomy

Another example of a navigational taxonomy can be found on the MyFlorida.com site map ([www.myflorida.com/taxonomy](http://www.myflorida.com/taxonomy)), where the top-level categories of the taxonomy, which also are the main navigation menu items, are Visitor, Floridian, Government, and Business. It is interesting to note that the file name for this site-map page has been named taxonomy.

Taxonomies can serve in the navigation of enterprise taxonomies. Enterprise taxonomies can be very large, but the top levels typically demonstrate some form of information organization for the enterprise. The purpose of this hierarchical organization is then not merely to retrieve documents but also to help users better understand the organization of the enterprise and its intranet and thus make better use of it.

It is open to dispute whether a website navigation structure can actually be called a “taxonomy,” even if it is well designed. A taxonomy, as a form of a controlled vocabulary, serves the purpose of indexing, classifying, or categorizing. It is a system in which a single term can be reused to index or classify more than one item, document, or webpage. In a website navigation structure, on the other hand, a menu label or site-map entry points/links to only a single webpage. This is a significant difference. The website navigation label needs to match or fit only a specific page. A taxonomy term, on the other hand, needs to be applicable to multiple, slightly different content items. Furthermore, website navigation labels are sometimes deliberately worded so as to intrigue or entice a user to visit a certain page within the site, rather than purely to inform or guide users. Finally, the hierarchical relationship and structure of a navigation structure reflect a custom experience, and make no attempt to follow standard conventions for hierarchical taxonomic relationships (explained in Chapter 4).

So, rather than considering every multilevel website navigation hierarchy as a kind of taxonomy, it would be more apt to look into how to make some taxonomies (that serve for indexing multiple

## **28 The Accidental Taxonomist**

content items) serve a second purpose of providing or influencing a website's hierarchical navigation structure, in cases when a unifying hierarchical structure is desired. This could be the case for enterprise taxonomies, especially if they are implemented both in a content management system and on an intranet and/or public website.

### **Taxonomies for License**

Although the primary objective of this book is to provide instruction on building taxonomies, it is not always necessary to build an entire taxonomy from scratch. Some or all of a taxonomy could be acquired from another source. While taxonomies for enterprises and content management systems should definitely be custom-created, a taxonomy for the indexing of documents or files in a given subject area could be purchased or licensed. Furthermore, taxonomies for license not only serve the purposes of indexing and content retrieval but may also provide an outline of a domain of knowledge. Many subject areas are already covered by existing published taxonomies. There are generic taxonomies for geographic places, industry types, product types, and so forth. In addition, lists of named entities are available from various sources. You might consider licensing an external taxonomy if the right taxonomy already exists and if creating one from scratch would be too great a task due to size, specialty subject area, and limited time. A licensed controlled vocabulary could be used for merely a single facet or for part of a larger set of taxonomies.

Taxonomies or controlled vocabularies that are available for license come from all kinds of sources: government agencies, professional associations, other nonprofit organizations, and a few commercial enterprises. Governmental published taxonomies available for license (or even free without a license) include LCSH, Library of Congress Thesaurus for Graphic Materials, MeSH, USDA National Agricultural Library Thesaurus, and the United Kingdom's Integrated Public Sector Vocabulary. The Getty Research Institute

(part of the J. Paul Getty Trust) is a reputable nonprofit provider of controlled vocabularies, including the Art & Architecture Thesaurus, Getty Thesaurus of Geographic Names, and Union List of Artist Names. The leading commercial vendor of prebuilt taxonomies, WAND Inc. ([www.wand.com](http://www.wand.com)), specializes in taxonomies in products and services. Some database publishers, such as Gale, a part of Cengage Learning, which created taxonomies initially as part of their own products, may also license these taxonomies alone (not a part of their content products).

The largest directory of taxonomies and thesauri available for use is Taxonomy Warehouse ([www.taxonomywarehouse.com](http://www.taxonomywarehouse.com)). The list is maintained by the taxonomy software vendor Synaptica. The database includes hundreds of taxonomies, including most of those mentioned previously. Some are simple controlled vocabularies or glossaries, but others are full-featured thesauri. Although some are hosted on the web, the data files (usually in CSV or XML formats) can be obtained for most of them. Figure 1.10 shows the information that Taxonomy Warehouse provides for a specific taxonomy. A single publisher may also offer numerous taxonomies on different subjects.

With the growing demand for tagging images, there are also an increasing number of “keyword lists” for photographers. Many of these are, in fact, hierarchical and faceted taxonomies. But since they usually do not include nonpreferred terms, they are rather simple taxonomies, despite the fact that some of them are very large (10,000 or 20,000 terms). Prices for these lists are relatively inexpensive compared with some of the commercial thesauri available, and some of the smaller keyword lists are free. A directory of these, called “Photography Keyword Lists & Free Keywording Resources for Lightroom,” is available on the website of one keyword-list vendor, Photo-Keywords ([www.photo-keywords.com/keywording-resources.php](http://www.photo-keywords.com/keywording-resources.php)), which is targeted especially at freelance professional photographers.

The screenshot shows the Taxonomy Warehouse interface. At the top left is a logo with a 'T' in a circle. To its right is the title 'Taxonomy Warehouse'. Below the title is a search bar with a dropdown menu set to 'Search All' and a 'Search' button. On the right side, there are four navigation links: 'Taxonomies', 'Events', 'Publishers', and 'Blogs \*', with 'Categories' listed below them. The main content area has a link 'Back to Search/Browse Results'. Below this is a section titled 'Title' containing the text 'USGS Thesaurus'. The next section is 'Details', which lists the following information: 'Taxonomy: USGS Thesaurus', 'Publisher: US Geological Survey', 'Vocabulary Type: Thesaurus', 'Description: The USGS Thesaurus is the principal controlled vocabulary supporting Science Topics. It is designed as a formal thesaurus conforming to ANSI/NISO Z39.19. The thesaurus is faceted, meaning its top terms delineate general aspects of information resources. The terminology is organized into 6 facets: Methods, Product types, Sciences, Topics, Time periods and Institutional structures and activities. The thesaurus can be viewed by selecting a hierarchical or keyword search.', 'More Info: <http://www.usgs.gov/science/about/>', 'Online At: <http://www.usgs.gov/science/about/term-simple.php?term=1>', 'Term Count: 1,001 - 5,000', 'Revision Cycle: Unknown', and 'Formats: Available in Electronic Formats'. The final section is 'Ontology', which lists: 'Published By [US Geological Survey](#) (Publishers)', 'Has Language [English](#) (Languages)', and 'Is About [Earth Sciences](#) (Categories)'.

Figure 1.10 Example display of information for a single taxonomy offered through the Taxonomy Warehouse

Formats may vary, but typically, taxonomies or thesauri that are made available for other uses are formatted in some form of XML, whereby all terms, relationships, nonpreferred terms, scope notes, and so forth are retained when they are imported into other taxonomy management systems. The use of XML and other interoperable taxonomy formats is described in greater detail in Chapter 11. The photography keyword lists are available in Excel or text files, in a format that allows them to be easily imported into photo management software, especially Adobe Lightroom, while maintaining the hierarchy.

If you acquire a taxonomy, however, you will likely want to modify or enhance it for your own needs, and in any case it will

require some maintenance over time. The following example case demonstrates how a generic taxonomy, taken as is, may not be ideal. A large-scale historical digitization project that coded early American election results used the Getty Thesaurus of Geographic Names. Even though the thesaurus includes historical place names, it was still found to be insufficient for the project's needs. It does not include all the towns and boroughs that were named in the elections project, and does not indicate exactly when various historical names were used or when boundaries were redrawn.

Licensing agreements may allow use of a taxonomy without a fee in some cases, but may prohibit for-profit use or require statements referring to the original copyright holder. If the taxonomy is treated as a published copyrighted work, whether free or for a fee, then there will also be restrictions on making changes to it.

A number of free controlled vocabularies, such as the Getty vocabularies, follow the model of Linked Open Data (LOD) and the Open Data Commons Attribution License, which allows modification of the vocabulary but require attribution to the original source. The Getty vocabularies website provides sample forms of attribution statements, such as: "This [title or report or article or dataset] contains information from Art & Architecture Thesaurus (AAT) which is made available under the ODC Attribution License."<sup>7</sup>

The policy for using and modifying MeSH is as follows:

If the use is not personal, (1) the U.S. National Library of Medicine must be identified as the creator, maintainer, and provider of the data; (2) the version of the data must be clearly stated by MeSH year, e.g., 1997 MeSH; and (3) if any modification is made in the content of the file, this must be stated, along with a description of the modifications.<sup>8</sup>



## **32 The Accidental Taxonomist**

Often you will want to make changes to the acquired taxonomy, so make sure the license permits changes. Also be aware that you are responsible for continued updating. Licensed taxonomies, both those that prohibit and those that permit changes, typically offer updates through an annual subscription. A solid understanding of how to create terms and relationships, as discussed in Chapters 3 and 4, is still necessary to manage prebuilt taxonomies. Therefore, acquiring a taxonomy from an external source does not eliminate the need for a taxonomist. Starting with a prebuilt taxonomy, though, can save time and make things much easier for the less experienced taxonomist. You can follow examples of term formats and relationships as you build out the taxonomy further.

### **History of Taxonomies**

Taxonomies are both new and old. “Both librarians and indexers were doing ‘taxonomy’ long before it became a hot topic in the 1990s,” wrote taxonomy trainer Jean Graef of the Montague Institute.<sup>9</sup>

### **Taxonomies in Cataloging and Indexing**

The earliest taxonomies were for classification, such as for organisms or for books, but each item could only go in one place in the taxonomy. For example, a book gets a single call number for its location on the shelf. In the field of library science, by the end of the 19th century, more practical taxonomies emerged that supported supplemental descriptive cataloging, which is not limited to one descriptive term per book. The leading controlled vocabularies for cataloging books have been the American Library Association Subject Headings (1895), Library of Congress Subject Headings (LCSH) (1898), and the Sears List, published originally as the List of Subject Headings for Small Public Libraries (1923).

These were simple controlled vocabularies lacking broader/narrower and related term relationships. LCSH used *See also* references for every kind of relationship and began to introduce broader term, narrower term, and related term references only in 1985.<sup>10</sup>

LCSH, back when it was still in its simpler form, was adopted by various periodical index publishers for the indexing of articles from multiple newspapers, magazines, and journals. These publishers included the H.W. Wilson Company (since acquired by EBSCO), which dated back to the start of LCSH, and Information Access Company (acquired by Gale, now a part of Cengage Learning) and ABI Inform (now ProQuest) in the 1970s. Cengage's and ProQuest's controlled vocabularies have diverged from LCSH over the years based on the work of their taxonomists.

Meanwhile, professional societies developed their own controlled vocabularies for indexing periodical literature in their fields since at least the early 1900s. These included the American Chemical Society's Chemical Abstracts Service founded in 1907. The word *thesaurus* was first used to refer to a controlled vocabulary for information retrieval purposes by Peter Luhn at IBM in 1957. Early published thesauri included the Department of Defense's *ASTIA Descriptors* in 1960 and the American Institute of Chemical Engineers' *Chemical Engineering Thesaurus* in 1961.<sup>11</sup> Standard thesaurus relationships emerged over time, and guidelines were developed that reinforced them, including UNESCO's 1967 guidelines, which formed the basis of the ISO 2788 standard of 1986<sup>12</sup> (superseded by ISO 25964 in 2011). Since the 1960s, various companies, government agencies, and professional associations have published dozens of specialized thesauri. In 1972, the new company Dialog began offering the first publicly available online research service, providing access to multiple bibliographic citation databases indexed with controlled vocabularies.

## 34 The Accidental Taxonomist

### Corporate Taxonomies

Up through the 1980s, however, taxonomy (thesaurus) development was mostly limited to large index or literature-retrieval database publishers, a few large companies, especially in the sciences (such as DuPont), and government agencies. The companies and government agencies that developed taxonomies did so mostly within specific subject areas. Taxonomies for an entire organization, that is, enterprise-wide taxonomies, first began to appear in the late 1970s, but their adoption was limited. According to taxonomy and knowledge management consultant Lynda Moulton, it was not so much a lack of interest but simply the limitations of software tools at the time that hindered a wider adoption of enterprise taxonomies. Moulton recalls teaching a number of thesaurus construction workshops during 1982–1984, attended by librarians and indexers from such companies as Liberty Mutual, John Hancock, Fidelity, MITRE, and Digital Equipment Corp.<sup>13</sup>

Contemporary library automation began to emerge in the late 1970s and systems for “special libraries” (corporate libraries and information management) as early as 1980. Although dedicated taxonomy management systems had not yet appeared on the market, these earlier systems included taxonomy management features. These included BiblioTech by Comstow Information Services (acquired by Inmagic in 1999, and since 2014 part of Lucidea), which was first installed at Polaroid in 1981, and TechLib, released in 1984, which was built on BASIS (acquired by OpenText in 1998). Comstow held a number of workshops that were devoted to thesaurus development for corporate libraries in the early 1980s.<sup>14</sup>

It was only in the late 1990s that a broader interest in taxonomies, and the corresponding tools to support them, developed. For example, the taxonomy consultancy Earley Information Science started working on classification, categorization, and metadata projects (essentially taxonomy, but not called that yet) to help their

clients make the most out of the Lotus Notes application, by building classification structures, forms, and navigation. In 1998, IBM introduced its Lotus Discovery Service, which “really called out the need for a taxonomy,” according to Seth Earley, so he and other consultants at the time provided services in creating taxonomies for Lotus Notes.<sup>15</sup>

### **The Growth of Enterprise and Web Taxonomies**

The emergence and growth of the web in 1990s was a major contributing factor in the growing interest in taxonomies, for several reasons. The web enabled smaller publishers to offer online information services. Companies started developing intranets that quickly expanded in size and required better navigation and search. “With the growth of the internet, there was a lot of interest in building to improve search results,” explained Synapse Corp. (now Synaptica LLC) co-founder Trish Yancey regarding the start of the company.<sup>16</sup> The proliferation of search engines, and then site search or enterprise search, also led to an interest in taxonomies as it became apparent that search alone was not sufficient. According to Jean Graef, “Taxonomy became hot when IT realized that search engines by themselves couldn’t solve the whole retrieval problem.”<sup>17</sup>

Finally, attention to site design and navigation through the new field of information architecture also put value on taxonomies. Indexer, information architect, and taxonomist Fred Leise wrote, “As the field of information architecture and the influence of Louis Rosenfeld’s and Peter Morville’s *Information Architecture for the World Wide Web* grew, the knowledge of library science-related information, such as faceted browsing classifications and the use of synonym rings as search improvements, spread more widely.”<sup>18</sup>

The growing interest in taxonomies in the 1980s and 1990s was also reflected in the growth of taxonomy management software. Software for creating and maintaining taxonomies was originally developed internally within the few large organizations that had

### 36 The Accidental Taxonomist

already developed taxonomies. In 1980 Comstow released Biblio-Tech, its fully integrated library system for corporate and government libraries, which included a module for thesaurus creation, fully integrated with the cataloging and indexing module. Battelle Columbus Laboratory released similar functionality in TechLib soon after.<sup>19</sup>

In the mid-1980s commercial PC software for thesaurus creation became available, including the desktop tools MultiTes, Term Tree, and TCS (later a part of WebChoir). Of these, only MultiTes is still offered today. Larger-scale client-server systems became available in the 1990s, reflecting the growing demand. Synapse Corp. had developed software to maintain taxonomies it was creating for others as a consulting service but soon found a market for the software itself and began selling the Synaptica taxonomy management system, the first commercial, web-based thesaurus management software, in 1999. Similarly, Access Innovations had been offering indexing services since 1978 but then found demand for its taxonomy management tool and has commercially offered its Data Harmony Thesaurus Master since 1998. Wordmap, another taxonomy software vendor, was founded in 1998. Content management systems and enterprise search solutions, which only really entered the market in the 1990s, have also begun to offer taxonomy management components or features.

The 1990s also saw the establishment of commercial vendors of taxonomies, including Synapse Corp. and WAND, both of which were founded in 1995, and the automatic taxonomy generator company Intellisophic in 1999.

The rise of the term *taxonomy* paralleled this growing interest in taxonomies. Former taxonomy consultant Ron Daniel got his start in the field working for the Department of Energy on its thesaurus. He recounts how, around 1997, it was starting to use the word *taxonomy* interchangeably with *thesaurus* and another term that hasn't become quite as popular, *synonymy*.<sup>20</sup> Earley recalls starting

to use the word *taxonomy* with clients around 1996 or 1997. Moulton recalls the adoption of the term *taxonomy* as follows:

Throughout my professional career, first as a technical librarian, then as a software developer and consultant, the operative terminology for my work was thesaurus.

... I first heard the term taxonomy applied to “organization maps,” in the early 1990s. ... In the late 1990s I began to see the term “taxonomy” routinely used to describe “terminology maps,” “topical hierarchies,” and “terminology relationships.” Before long, taxonomy became the de facto label for topical navigation schemes on commercial websites that had a focus on text content retrieval. ... At some point I recognized that the term “thesaurus” was not understood by IT and business management professionals. So, about 2000, I adopted taxonomy to cover any controlled vocabulary being developed or applied in any indexing, metadata management or retrieval situation. ... To this day, I use thesaurus and taxonomy interchangeably depending on which word will most likely resonate with my audience.<sup>21</sup>

Our online survey completed by 148 taxonomists in May 2015 also confirmed the relatively later adoption of the term *taxonomy*. Whereas 26 (17.6 percent) of the respondents had been involved in taxonomy work as we define it (taxonomies, controlled vocabularies, metadata for classification or tagging, thesauri, or authority files) for more than 15 years, only 11 (7.5 percent of the total) reported that their work was specifically called “taxonomy” as long ago as 2000 (15 years prior to the survey). Going back 20 years or more (1995 or earlier), 18 respondents, or 12.2 percent, claim to have already been in the field, but only 2 of them said it was called “taxonomy” then. (See Appendix A, Questions 2 and 3.)

### 38 The Accidental Taxonomist

Another way to track the growth in the popularity of taxonomies is to track the mention of taxonomies in periodical literature, especially in business and trade journals. While many of these articles may be about specific subject taxonomies, rather than information taxonomies in general, searching on the plural word *taxonomies*, rather than *taxonomy*, focuses the results more on the creation of generic information taxonomies. Looking at Gale's Business Collection subscription database on InfoTrac for the word *taxonomies* in titles and full text shows a marked increase from 1998 to 2002 and then a steadier occurrence in subsequent years, except for a spike in 2010, as shown in Table 1.1. Business Index covers 3000–4000 magazines, journals, newsletters, industry reports, and newswires.<sup>22</sup> "Taxonomies" is not a subject term in the Gale databases, but rather the slightly broader "controlled vocabularies" is used.

The turning point came around 2000. In the summary of the European Business Information Conference (EBIC) conference in 2000, Tom Koulopoulos, president of the Delphi Group and renowned writer and public speaker on knowledge management, declared, "Taxonomies are chic." Since then taxonomies have been a popular topic in conference presentations and workshops. The Montague Institute held its first taxonomy roundtable in 2000. A significant number of taxonomies had become available publicly (usually for licensing), so in 2001 Synapse Corp. (now Synaptica, LLC) launched its Taxonomy Warehouse website directory of taxonomies. Taxonomy consultant Marcia Morante recalled as follows:

The year 2000 was probably the very beginning of the commercial taxonomy wave. That was the year that I started with Sageware, and we still had to do a lot of explanation. But by that time, there were definitely a few companies whose business was built around taxonomies.<sup>23</sup>

Table 1.1 Number of periodical articles about taxonomies in Gale’s Business Collection database on InfoTrac

Year	“taxonomies” in article titles	“taxonomies” in article text	“controlled vocabularies” Subject
1998	0	6	0
1999	0	16	0
2000	3	92	0
2001	6	111	0
2002	8	231	0
2003	8	180	0
2004	14	208	0
2005	7	204	0
2006	5	200	1
2007	7	206	9
2008	4	214	8
2009	6	217	10
2010	10	365	15
2011	8	159	6
2012	6	192	5
2013	3	169	4
2014	4	206	31

Although newer buzzwords, such as *folksonomy*, *social networking*, and *Web 2.0*, superseded *taxonomy* in their usage in the 2000s, a sustained interest in taxonomy and taxonomists continues. Search industry analyst Steve Arnold analyzed web traffic on Google from 2002 to 2008 on the term *taxonomy* and found it continuing to remain strong, stronger than *CMS* (content management systems). He concluded that “taxonomy is a specialist concept that seems to be moving into the mainstream.”<sup>24</sup>



## Endnotes

1. Gale Hodge, *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files* (Washington, DC: The Digital Library Federation Council on Library and Information Resources, 2000), 1, accessed December 20, 2015, [www.clir.org/pubs/reports/pub91/pub91.pdf](http://www.clir.org/pubs/reports/pub91/pub91.pdf)
2. *Ibid.*, 4–7.
3. National Institute of Standards Organization, ANSI/NISO Z39.19-2005 (R2010) *Guidelines for Construction, Format, and Management of Monolingual Controlled Vocabularies* (Bethesda, MD: NISO Press, 2010), 166.
4. Alice Redmond-Neal and Marjorie M. K. Hlava, eds., *ASIS&T Thesaurus of Information Science, Technology, and Librarianship*, 3rd ed. (Medford, NJ: Information Today, 2005).
5. Tom Gruber, “Ontology,” accessed December 20, 2015, [tomgruber.org/writing/ontology-definition-2007.htm](http://tomgruber.org/writing/ontology-definition-2007.htm)
6. This image, reprinted with permission of the authors, first appeared in Murat Sensoy and Pinar Yolum, “Ontology-Based Service Representation and Selection,” *IEEE Transactions on Knowledge and Data Engineering* 19, no. 8 (2007). It is also available at [mas.cmpe.boun.edu.tr/project/AgentBasedSemanticWebServices.htm](http://mas.cmpe.boun.edu.tr/project/AgentBasedSemanticWebServices.htm)
7. Patricia A. Harpring (Managing Editor of the Getty Vocabulary Program, Getty Research Institute), email to author, January 13, 2016.
8. MeSH Memorandum of Understanding, accessed December 20, 2015, [www.nlm.nih.gov/mesh/termscon.html](http://www.nlm.nih.gov/mesh/termscon.html)
9. Jean Graef, email to author, November 21, 2008.
10. Alva Stone, “The LCSH: A Brief History of the Library of Congress Subject Headings, and Introduction to the Centennial Essays,” *Cataloging & Classification Quarterly* 29, no. 1–2 (2000): 1.
11. Jean Aitchison and Stella Dextre Clarke, “The Thesaurus: A Historical Viewpoint with a Look to the Future,” in *The Thesaurus: Review, Renaissance, and Revision*, eds. Sandra K. Roe and Alan R. Thomas (Binghamton, NY: Haworth Press Inc., 2004), 7.
12. *Ibid.*, 8.
13. Lynda Moulton, telephone interview with the author, October 19, 2009.
14. Lynda Moulton, email to author, October 19, 2009.
15. Seth Earley, telephone interview with author, November 22, 2008.

16. Kimberly S. Johnson, "International Information Provider Buys Franktown, Colo., Taxonomy Company," *Denver Post*, June 30, 2005.
17. Jean Graef, email to author, November 21, 2008.
18. Fred Leise, email to author, December 2, 2008.
19. Lynda Moulton, email to author, October 19, 2009.
20. Ron Daniel, telephone interview with author, December 1, 2008.
21. Lynda Moulton, email to author, November 9, 2009.
22. Gale/Cengage Learning Database Title Lists, [solutions.cengage.com/Gale/Database-Title-Lists](http://solutions.cengage.com/Gale/Database-Title-Lists), updated June 30, 2015. A count of titles without an end date. Titles are added and dropped over the years, so an exact title count for a span of years is not possible.
23. Marcia Morante, email to author, November 21, 2008.
24. Steve Arnold, "Taxonomy: Silver Bullet or Shallow Puddle," *Beyond Search* blog, September 27, 2008, accessed December 20, 2015, [arnoldit.com/wordpress/2008/09/27/taxonomy-silver-bullet-or-shallow-puddle](http://arnoldit.com/wordpress/2008/09/27/taxonomy-silver-bullet-or-shallow-puddle)

## About the Author



Heather Hedden has been active in developing and editing taxonomies since 1995. She has worked as an independent consultant (Hedden Information Management, [www.hedden-information.com](http://www.hedden-information.com)) and currently is a senior vocabulary editor at Cengage Learning. Heather is also an instructor of taxonomy development through the continuing education program of Simmons College School of Library and Information Science.

Heather began her taxonomy work as a controlled vocabulary editor at Cengage Learning's predecessor companies, Information Access Company, Gale Group, and Thomson Learning. Her work there included converting the controlled vocabularies into ANSI/NISO standard thesauri, updating and adding new taxonomy terms, communicating new terms and policies to indexers, mapping taxonomies, and creating new user interface taxonomies. Heather then worked as the information taxonomist at an enterprise search software startup, Viziant Corporation, where she was solely responsible for developing all taxonomies used with machine-learning-based auto-categorization. After that, she worked as the taxonomy manager at First Wind, where she developed taxonomies for search and navigation in the company's SharePoint intranet. Heather worked as a taxonomy consultant employee of Project Performance Company and then pursued full-time independent taxonomy

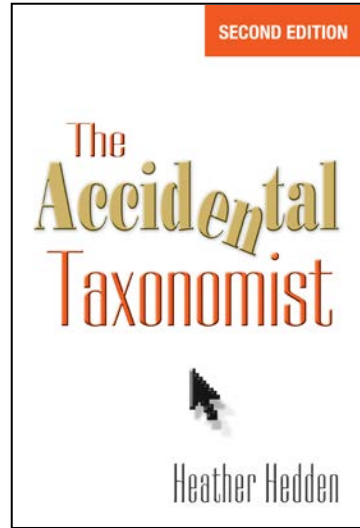
consulting through Hedden Information Management. Over the years, Heather has also done freelance back-of-the-book indexing.

Heather is the author of *Indexing Specialties: Web Sites* (2007), the chapter “Controlled Vocabularies, Thesauri, and Taxonomies” in *Index It Right! Advice From the Experts* (Vol. 2, 2010), and the chapter “Indexing Arabic Names” in *Indexing Names* (2012), all of which are published by Information Today, Inc., on behalf of the American Society for Indexing (ASI). She has also published numerous articles in journals, such as *EContent*, *Intranets*, *Computers in Libraries*, *Journal of Digital Asset Management*, *The Indexer*, and *Key Words* (bulletin of ASI). Heather has given presentations and workshops nationally and internationally at conferences, including Taxonomy Boot Camp, Enterprise Search Summit, Gilbane Conference, Content Management Professionals, SLA, ASI, Indexing Society of Canada, Society of Indexers (UK), and Netherlands Society of Indexers.

Her professional association memberships include Special Libraries Association and its Taxonomy Division, the Information Architecture Institute, and ASI. Heather has served as president of the New England Chapter of ASI, manager of the Web Indexing special interest group of ASI, and founder/manager of the Taxonomies & Controlled Vocabularies special interest group of ASI. She is currently (2015–2018) a member of the board of the American Society for Indexing and a member (2015–2016) of the NISO working group “Development of Standards to Support Bibliographic Data Exchange.”

Heather has a BA from Cornell University and an MA from Princeton University. She lives with her husband in Carlisle, Massachusetts.

If you enjoyed reading this chapter of *The Accidental Taxonomist*, Second Edition, you can order it from the following online retailers.



**amazon.com**<sup>®</sup>

 **Information Today, Inc.**