Taxonomy Tool

Heather Hedden

There are quite a few products on the market called taxonomy tools. To choose just a few for a comparative evaluation requires narrowing the field a bit. Since taxonomies are typically used for tagging and searching, a number of taxonomy products integrate taxonomy creation and maintenance features along with capabilities such as auto-categorization, human tagging, or search to provide more complex, full-featured solutions. This article, however, looks at products that facilitate only taxonomy creation and maintenance.

Taxonomy tools are used by individuals who call themselves taxonomists or who otherwise are familiar with best practices for creating taxonomies, including librarians, controlled vocabulary editors, lexicographers, and some information architects. The tools are not difficult to use, but creating a good taxonomy requires specific skills and knowledge of categorization practices. The use of the software alone does not necessarily result in a usable taxonomy, just as a good HTML editing tool does not necessarily result in a good website if the user is not skilled in the techniques of web design.

There are different definitions and types of taxonomies: simple lists of terms with synonyms, hierarchical trees, faceted categories, standard thesauri, and complex ontologies. The products reviewed here serve the creation of all of these types of taxonomies, except for limitations in ontologies. The designation "taxonomy" can refer to all of these or, more specifically, to a hierarchical, tree-type taxonomy. We will refer to taxonomies in the broader sense. An information retrieval thesaurus can be considered a more complex type of taxonomy, which supports not only hierarchical relationships but also associated term relationships, cross-references from nonpreferred (used for) terms. The products evaluated here each support these thesaurus characteristics, so they are marketed as "thesaurus construction tools," rather than taxonomy tools. Yet thesaurus tools are the best options for creating a simple hierarchical taxonomy as well.

The basic requirement of a taxonomy tool is to maintain terms and their associated relationships and other attributes. The relationships are reciprocal between pairs of terms, so by using a taxonomy tool, the user only needs to create or edit the relationship in one place. If the user decides to rename or delete a term, all its relationships will reflect the change. Support for optional scope notes and user-defined classification categories for each term is also standard in a taxonomy tool.

The products covered in this evaluation all meet the basic requirements and share several additional features. These include designating candidate and approved terms, indicating term creation date and modification date, permitting multiple hierarchies (polyhierarchies), and disallowing illegal term relationships. They all run only on Windows and include online Help. They are available as affordable single-user desktop tools in addition to having multi-user versions and can export taxonomies in platform-neutral formats for use in other systems.

Product name: MultiTes Pro, version 2007.02.01
Product vendor: Multisystems (Miami)
Price: $295 single user; $1295 for 5 users; $2495 for 10 users; $3950 enterprise deployment
*www.multites.com*

TAXONOMY DISPLAY

A thesaurus is typically displayed alphabetically, with relationships and attributes listed at each term, whereas a simple taxonomy, with its emphasis broader-narrower (parent-child) relationships, is typically displayed as a hierarchy. Therefore, a thesaurus and taxonomy tool could present the list of terms more than one way. MultiTes displays the taxonomy in an alphabetical list of all terms. To view a hierarchy of the taxonomy requires selecting the "Hierarchical" or "Top term" options from the Report menu, which generate simple text files to the screen.

TERM DISPLAY AND EDITING

Creating and editing terms and their relationships is easy and efficient in MultiTes. To view or edit a term's details involves clicking on the desired term and calling up a new pop-up window. The term window provides the term name, classification, relationships, notes, and other attributes, which are logically grouped. From here the details window of any associated term can also be jumped to, a feature limited in the other tools.

MultiTes has a clean, compact term-editing interface. All types

Raundup
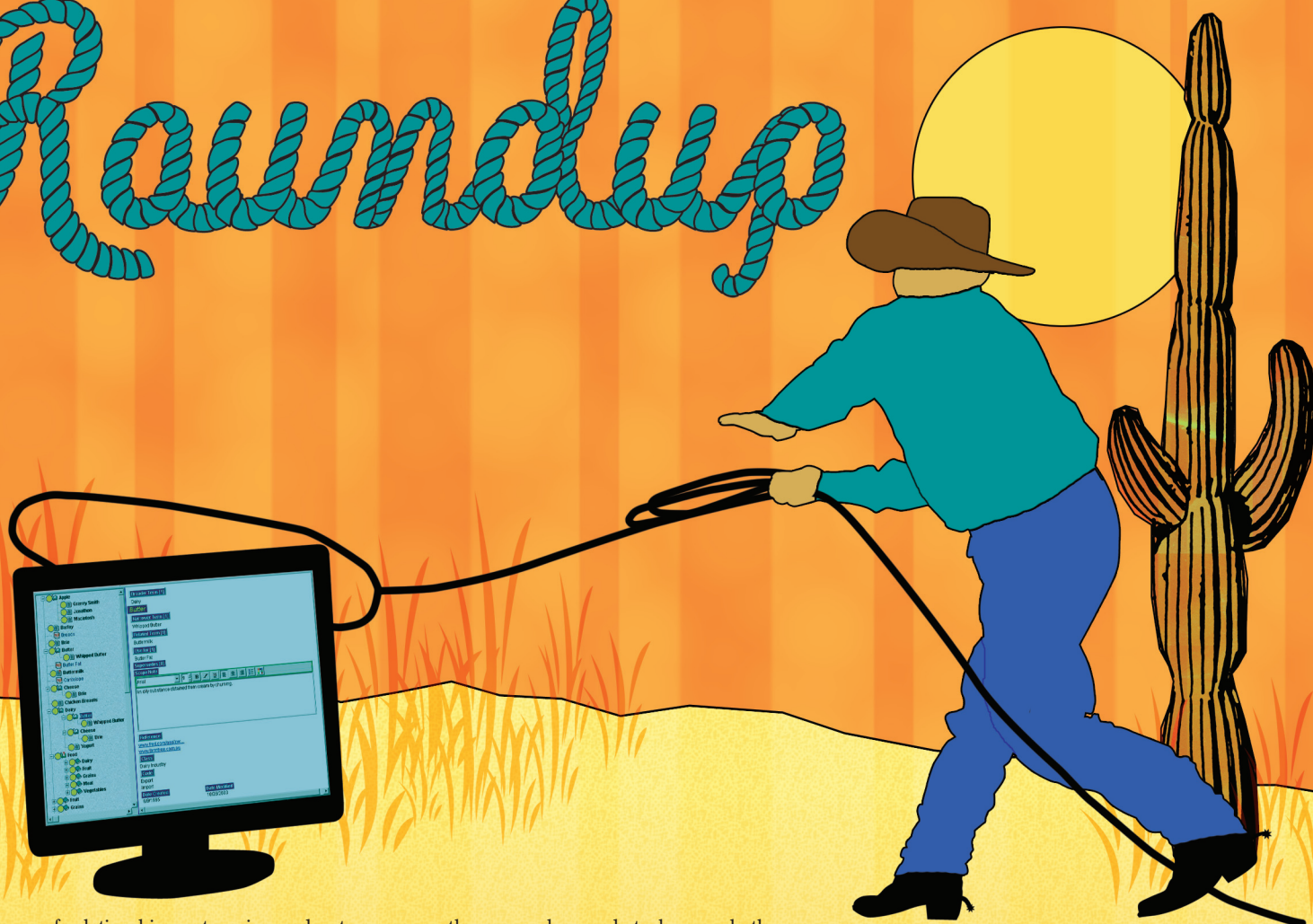
of relationships, categories, and notes can be added by selecting their label code from a single alphabetical drop-down list. The main term-editing drawback is that existing relationships cannot be changed in one step, but rather they have to be deleted and new ones added.

MultiTes allows the creation of relationships to existing terms and the creation of relationships to new terms that the user simultaneously creates. When creating a relationship to a term that does not yet exist in the taxonomy, a pop-up window alerts the user and asks if it should proceed with creating the new term. This is a nice feature, because often it is difficult to remember if the target term of a link already exists.

TAXONOMY SEARCHING

Taxonomy navigation is important for the user, who needs to know whether a given term has already been created. In MultiTes, the user simply types the start of the term, with truncation within a word, and hits enter. In addition, an Advanced Search option allows the user to select from terms which start with truncated text, contain text, are flagged, belong to a category, have a status, have a term number, whose date contains text, or whose note contains text. A list of matching terms then appears in the search window. From the list one can jump to an individual record. Another option is copying the list of terms to the clipboard for offline manipulation.

USER-DEFINED RELATIONSHIPS AND ATTRIBUTES

The ability to define relationships, types of notes, and categories of terms is an important set of features for making a taxonomy tool versatile and extensible. More complex, or semantic, types of relationships (such as "produced by," "owned by," and "utilized for") are what can distinguish an ordinary taxonomy from an ontology. MultiTes allows unlimited user definition of relationships, as long as they fall into one of three main categories: hierarchical (broader/narrower or parent/child), associative (related terms), or equivalent (use/use for). The creation of reciprocal terms is also enforced.

Similarly, MultiTes supports user-defined categories and notes for terms. Categories are often used in taxonomies to classify terms by source, for end use, or for any purpose the taxonomy developer may have. User-defined notes might be

desired for different audiences: taxonomists, indexers, and the end users. MultiTes does not support additional user-defined attributes, but there is no reason why the notes field cannot be used for any purpose, such as a physical address for a company name term.

IMPORTING, EXPORTING, AND REPORTS

MultiTes accepts imported data as a text file, with specifications of one or more relationships or attributes for the terms. The data must follow a specific format but does not use tags. This is the only import format that MultiTes accepts. MultiTes has the advantage over the other tools of accepting imported data not merely as files but also simply as pasted text from a clipboard. This is often a practical way for the user to batch load even just a couple of terms or relationships.

The features of exporting and generating reports are combined in MultiTes, allowing a full range of displays, subsets of terms to include or exclude, and term information to include or exclude. This is in addition to the different output formats. The output formats available are ANSI text, HTML (as a webpage), XML, and CSV delimited, or an output directly

to the printer. There is a full range of display formats and choices for inclusion/exclusion.

Product name: Term Tree 2000, version 2.3
Product vendor: This to That Pty., Ltd. t/a A.C.S. Active Classification Solutions (Australia)
Price: $800 Australian (about $700)
www.termtree.com.au

TAXONOMY DISPLAY

Term Tree's user interface features a split screen with the browsable taxonomy

in the left pane and the editable relationships and details of a selected taxonomy term in the right pane.

The terms are listed alphabetically but with the option to expand to show narrower terms, which is lacking in MultiTes. Nonpreferred terms are interspersed among the preferred terms. The display makes use of various graphical icons next to all terms. Another nice feature is the View menu, which allows the user to limit the taxonomy display to just top terms, preferred terms, nonpreferred terms, unauthorized terms, and so on, which in other products can only be done as a report.

TERM DISPLAY AND EDITING

Although Term Tree's interface is somewhat graphically cluttered, the tool offers an impressive number of options for term creation, linking, and editing. Term details can be edited either from the term detail pane or by clicking on a term in the taxonomy list and using a context menu, toolbar buttons, or even keyboard shortcuts (function keys, insert, delete, and so on). Finally, relationships can be created by dragging a term in the left pane taxonomy list and dropping it
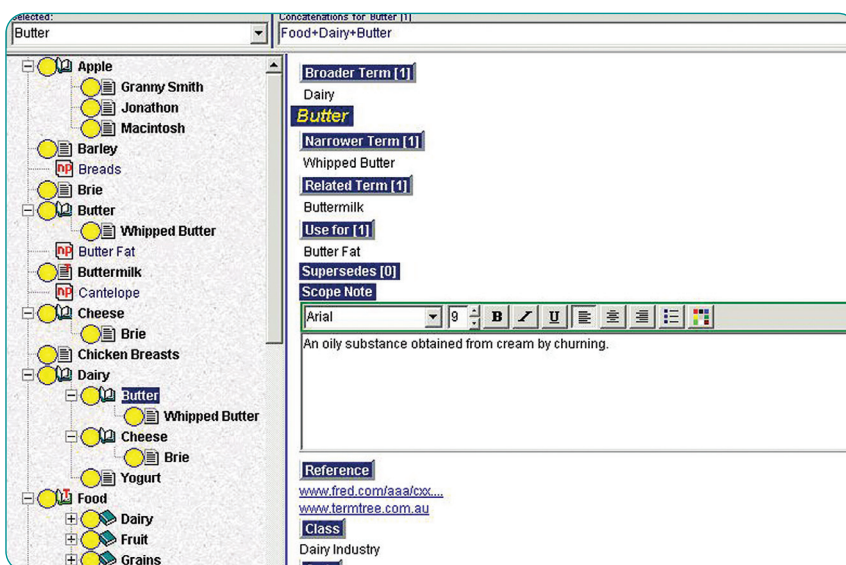
into the relationship type of another term in the right pane details screen.

In addition to the standard relationship types, Term Tree also has a "superseded term" relationship for when a term is deactivated and superseded by another term. New terms can be created by starting from an existing term and creating a relationship or from scratch by various methods. Term Tree spell checks new terms and scope notes as they are typed in. A drawback, however, is that editing the term name itself, such as changing its wording, cannot be done without deleting the term and then adding back the new wording of the term.

TAXONOMY SEARCHING

Term Tree supports searching the start of terms in the taxonomy with type-ahead scrolling.

An additional search option permits searching other fields, such as scope note, and allows restricting by source, code, class, and date. If multiple terms match the search, rather than getting a list of matches, the user needs to select "next occurrence" from the menu. In my opinion, a list of matches would be more practical.

USER-DEFINED RELATIONSHIPS AND ATTRIBUTES

In Term Tree—although system definitions permit the user to rename the relationship labels of Use, Use for, Broader Term, Scope Note, and so on—creating additional kinds of relationships or types of notes is not supported. Actually, the system provides two note types called Scope Note and User Note, so having the possibility for a second kind of note is, in most cases, sufficient. The user can, however, create and name any number of categories, called Classes in Term Tree. There is even an additional type of category called Codes, although the need for a distinction between Classes and Codes is not obvious.

IMPORTING, EXPORTING, AND REPORTS

Term Tree supports importing data as a tagged ASCII text file and in the proprietary formats of MultiTes and A.K.A., but not XML. Of the tools tested, Term Tree can export to the greatest number of file formats: XML, CSV comma delimited, tab delimited, Excel spreadsheet direct, HTML (as a website), Windows Explorer, Term Tree's own tag format, and the proprietary netcat and MetaBrowser formats.

Reports can be generated from Term Tree in more types than most users need for various thesaurus and records management outputs. The report output format, however, is limited to MS Word or direct to printer. A full range of term inclusion/exclusion is possible, as in MultiTes.

Product name: TCS-10, version 2.26
Product vendor: Webchoir, Inc. (Los Angeles)
Price: $399 Personal edition (single user); $1495 for 10 users
*www.webchoir.com*

TAXONOMY DISPLAY

Webchoir's TCS-10 offers two views between which the user can toggle: an alphabetical list and a true hierarchy. The hierarchical view lists only the top hierarchical terms, which can be expanded to show narrower terms. The "expand all" and "collapse all" buttons are nice features. TCS-10 is the only tool of the three that does not include nonpreferred terms in the taxonomy list. Since nonpreferred terms are still searchable in TCS-10, this is fine, and the exclusion indeed makes the term list easier to browse.

TCS-10 has the benefit of supporting the creation of multiple hierarchies within one thesaurus project. This way multiple authority files or facets of the same taxonomy project can be created and interrelated. The user can alternate between them using the Window option on the menu.

TERM DISPLAY AND EDITING

Like Term Tree, TCS-10's user interface features a split screen with the browsable taxonomy in the left pane and the details of a selected term in the right pane. A third section at the bottom of the taxonomy pane allows single-click online searching of a selected term through a chosen web search engine. TCS-10 differs by making a distinction within its right pane between a mere display and an editable window. When a term is selected the

The TCS-10 taxonomy display with selected term details for viewing, not editing
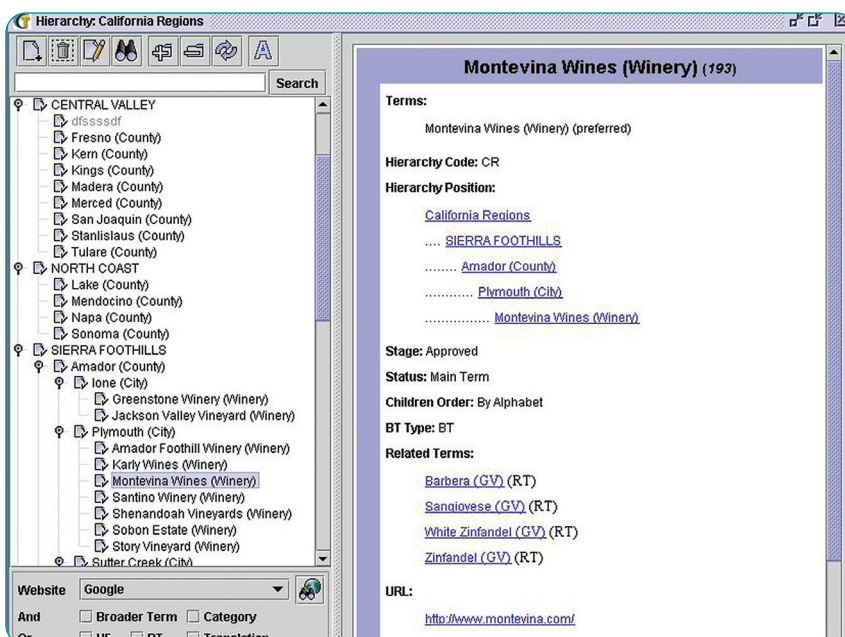
All three tools have comparable features, and no one is clearly superior. A choice should be made according to user needs and preferences. However, this chart compares some of the features that do vary between these products.

Taxonomy Tool Feature Comparison

Feature	MultiTes	Term Tree	TCS-10
User-definable relationships	Y	N	P
User-definable notes	Y	N	Y
User-definable attributes	P	N	Y
Spell checking	N	Y	N
Single-step new term & relationship creation	Y	Y	N
Drag & drop relationship adding	N	Y	P
Descriptor name editing	Y	N	Y
XML importing	N	N	Y
Multilingual support	Y	N	Y
Single-step move branch	N	N	Y
Expandable hierarchy display	N	P	Y
Online Help usability	Y	Y	N
Note: (Y = yes, N = no, P = partial)			

uneditable display appears first, and the edit command must be selected in order to make changes or additions, and afterwards one must click "Close." This added step can slow down work. The editable fields for each term additionally include URLs and images, but what is blatantly missing are broader term and narrower term relationships. This is because terms can be created only as narrower terms of existing terms. While this might be the most common method by which taxonomists work, it should not be the only method permitted.

TCS-10 does have a limited drag-and-drop feature, which enables moving only a term from one broader term to a different broader term. This is done within the taxonomy tree display pane, though, which necessitates collapsing hierarchies so that both the moved term and the target broader term appear within the same display window.

TAXONOMY SEARCHING

As with MultiTes and Term Tree, TCS-10 features both a search box at the top of the scrollable taxonomy list and also an advanced search option. Although it does not support truncation, advanced search will match any word within a term, not merely the first word. As this kind of search can yield multiple results, a window pops up with all matching results from which any matched term can be selected, a feature found only in the advanced search of MultiTes. TCS-10's advanced search does not have as many limiting options as advanced search in the other tools, but it does has the advantage of offering guided Boolean search.

USER-DEFINED RELATIONSHIPS AND ATTRIBUTES

TCS-10 supports unrestricted user-defined variations of the "use" cross-reference type for relationships and related term relationships, but it restricts the broader/narrower relationships. Instead, it provides the option to choose from three additional types of broader/narrower relationships in accordance with thesaurus standards: BTG (Broader Term—Generic), BTI (Broader Term—Instance), and BTP (Broader Term—Partitive), but only one BT relationship type of each kind is allowed.

TCS also supports unlimited user-defined notes and other attributes. User-defined attributes, ambiguously called "User Relation Type," could be used for location or contact information for a named entity term. A unique feature to TCS-10 is that one can even create authority control for a relation type, such as a look-up table of state names.

IMPORTING, EXPORTING, AND REPORTS

Of the tools tested, TCS-10 supports the greatest number of formats for importing data, which include ASCII, XML, and MARC. For the import of ASCII text files there are additional submenu options and an ASCII import wizard. For importing XML, a tag-mapping table needs to be completed. TCS-10 supports exporting XML, ASCII, MARC, HTML (as a website), and a proprietary format for Webchoir's other indexing and search products. User-created look-up tables can also be exported.

Report options from TCS-10 are comparable with the other taxonomy tools, and terms can be included/excluded by various criteria. Additional kinds of reports include URL list, image list, and category list. The reports are output as either ASCII text files or HTML (without hypertext). ■

HEATHER HEDDEN (HEATHER@HEDDEN.NET) IS A TAXONOMIST AT VIZIANT CORP., A TEACHER OF CONTINUING EDUCATION WORKSHOPS AT SIMMONS COLLEGE GRADUATE SCHOOL OF LIBRARY AND INFORMATION SCIENCE, AND FOUNDER AND MANGER OF THE TAXONOMIES & CONTROLLED VOCABULARIES SIG OF THE AMERICAN SOCIETY OF INDEXERS.
COMMENTS? EMAIL LETTERS TO THE EDITOR TO ECLETTERS@INFOTODAY.COM.