

Comparative evaluation of thesaurus creation software

Heather Hedden

This article provides a comparative evaluation of three software programs used for the creation and editing of controlled vocabularies, thesauri, or taxonomies: MultiTes, Term Tree 2000, and WebChoir TCS-10. These thesaurus tools are comparable in their features and are all available in single-user desktop versions affordable for the freelance taxonomist.

Indexers have the skills, and sometimes the need, to go beyond creating indexes and create controlled vocabularies or thesauri used for indexing. But understanding the principles of thesaurus creation is not enough for one to efficiently complete a project. As in the field of indexing, having a good software tool is also important.

The size and scope of a thesaurus project can vary greatly. It could be a simple term list that an individual indexer creates to help with the indexing of a longer than average book. At the other extreme the project could involve a large collection of inter-related authority files of tens of thousands of terms managed by multiple controlled vocabulary editors in a large organization. Similarly, the software tools available vary in their capabilities, number of users supported, and price. At the high end, a number of thesaurus products (also called taxonomy tools) integrate thesaurus creation and maintenance features along with the capabilities of auto-categorization, human indexing, and/or search to provide more complex full-featured solutions. Examples include Early & Associate's Wordmap, Data Harmony's MAIstro (which includes both Thesaurus Master and the indexing tool M.A.I.), and Dow Jones' Synaptica. This article, however, looks at products that facilitate only thesaurus creation and maintenance and have affordable single-user versions.

A thesaurus can be considered as a type of taxonomy which supports not only hierarchical relationships but also associated term relationships, cross-references from nonpreferred ('used for') terms, and the additional option of notes for each term. ('Controlled vocabularies, thesauri, and taxonomies' (Hedden, 2007) explains the definitions of thesaurus and taxonomy in greater detail.) Published national and international standards provide specific guidelines for the creation of information retrieval thesauri. The products evaluated here all support the characteristics of a thesaurus in accordance with these standards. They are indeed marketed as 'thesaurus construction tools,' although they are sometimes also called taxonomy tools. For simplicity, though, we will refer to the person using the software to create a thesaurus as the 'taxonomist' or simply the 'user.'

The basic requirement of a thesaurus tool is to maintain terms and their associated relationships and other attributes. All these relationships are reciprocal between pairs of terms. By using a thesaurus tool, the taxonomist

only needs to create or edit the relationship in one place. If the user decides to rename or delete a term, the change is reflected in all its relationships. Merging and subsuming terms may also be supported. Features for optional scope notes and user-defined classification categories for each term are also expected in a thesaurus tool.

The products covered in this evaluation all meet the above basic requirements and share several additional features. These include designating candidate and approved terms, indicating term creation date and modification date, permitting a term to belong to multiple hierarchies (polyhierarchies), and disallowing illegal relationships (e.g. circular relationships). They all run only on Windows and include online Help. They can also export taxonomies in platform-neutral formats for use in other systems. Free demo versions for each are available for download.

The following topics will be evaluated and compared for each product: thesaurus display, term display and editing, thesaurus searching, user-defined relationships and attributes, rules enforcement, importing and exporting or reports, and online help and tutorials.

Thesaurus display

A thesaurus is typically displayed alphabetically, with relationships and attributes listed at each term. A simple taxonomy, on the other hand, with the emphasis on broader-narrower (parent-child) relationships is typically displayed as a hierarchy. Thus, a thesaurus/ taxonomy tool could conceivably present the list of terms more than one way, and indeed the different tools do offer different displays. Thesaurus tools also differ in whether they create single isolated thesaurus files or named hierarchies that can be related to each other.

Term display and editing

Each thesaurus term has relationships and other details that can individually be displayed and edited, although the various thesaurus tools differ in how this is done and in what attributes are supported. Creating terms and their relationships can be performed through various means: main menu selections, context menu selections, shortcut key

combinations, other keyboard commands, toolbar buttons, and mouse drag and drop.

Thesaurus searching

Thesaurus navigation is important not merely for the end-user of the implemented thesaurus, but also for the creator of the thesaurus, who needs to know, for example, whether a given term has already been created.

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 thesaurus tool versatile and extensible. Although standard thesaurus relationships are limited to broader term, narrower term, related term, and use/used for, more specific user-defined relationships might be desired. More complex types of relationships are what distinguish an ordinary thesaurus from what is called an ontology. Relationship types between terms that give more meaning, such as 'produced by,' 'owned by,' 'purchased by,' and 'utilized for,' are known as semantic relationships. With the growing interest in the Semantic Web, it is becoming increasingly important that a thesaurus tool also support user-defined relationships.

Categories are often used in taxonomies to classify terms for end-use, by source, or for any purpose the thesaurus developer may have. Categorizing terms makes them easier to batch edit for the taxonomist and also possible to designate for certain end-use search interface characteristics. Such user-defined categories could also be used to distinguish generic terms from named entities, terms that the user decides may be permitted to have certain semantic relationships, and terms that belong within a facet for a faceted taxonomy. There could be a master taxonomy, for example, and certain subsections categorized for certain audiences or services.

Rules enforcement

A thesaurus should follow certain rules, such as not having terms that duplicate within the same hierarchy or category, terms that lack hierarchical relationships with any other terms (orphan terms), or terms that have a broader term which is narrower than a narrower term of the first term (circular references). Not necessarily 'rules', but controls may also be desired over additional areas, such as whether to allow multiple 'use' references.

Importing, exporting, and reports

In addition to building a thesaurus from scratch, i.e. manually typing in each term, a taxonomist may want to take advantage of external lists of terms to import and incorporate into the thesaurus. Most often these might be lists of names, organizations, or places, but it could also be the incorporation of legacy taxonomies, along with their

relationships, into a new system. Thus, importing or batch loading of data into a thesaurus tool is an important feature.

Exporting the thesaurus into formats that can then be imported into other systems is crucial for a thesaurus tool. A thesaurus is not just to look at but is also to be used in the indexing/tagging of documents or web pages and then by a final end-user for search and retrieval of those documents and pages, which is done in one or more other software systems.

Generating interim reports of various kinds can aid the taxonomist in the task of building a thesaurus. Occasionally a thesaurus is published for third party use, in which case various outputs, including a printed document, might be desired.

MultiTes Pro

Product name: MultiTes Pro, version 2007.02.01

Product vendor: Multisystems (Miami, Florida, USA)

Price: US\$295 single user; US\$1,295 for 5 users; US\$2,495 for 10 users; US\$3,950 enterprise deployment.

www.multites.com

Thesaurus display

MultiTes displays the thesaurus in an alphabetical list of all terms, interspersing nonpreferred terms (in italics) among the preferred terms. There are no options to change the user interface display. To view a hierarchy of the thesaurus it is necessary to select the 'Hierarchical' or 'Top term' options from the Report menu, which can generate simple text files to the screen. The 'Hierarchical' report displays the full parent/child hierarchy path for all broader and narrower relationships for each term. The 'Top term' report displays a simply hierarchy of the path of narrower terms, if any, under each term.

MultiTes' display, while limited to alphabetical, does include some additional information for each term which the other tools do not include in their alphabetical displays. Making use of the full width of the screen in a tabular format, MultiTes includes columns for term status, types and number of relationships, note types, category names, and language. The user can customize the display to show only the desired columns of information. While it is a minor inconvenience to have to go through the steps of generating a report to obtain a hierarchical view in MultiTes, the user has the options of restricting a hierarchical display to a certain subsection or category of the entire thesaurus. While the hierarchy cannot be expanded in the main view, more terms fit into the MultiTes display, about 38, compared with about 28 in Term Tree and TCS-10.

MultiTes creates single thesaurus files with the extension .th2. Relationships cannot be created across files. To create multiple authority files or facets with interrelationships, these each have to be created within the same file and designated by user-defined categories.

Term display and editing

In MultiTes, since the thesaurus list takes up the entire window width, to view or edit a term's details involves clicking

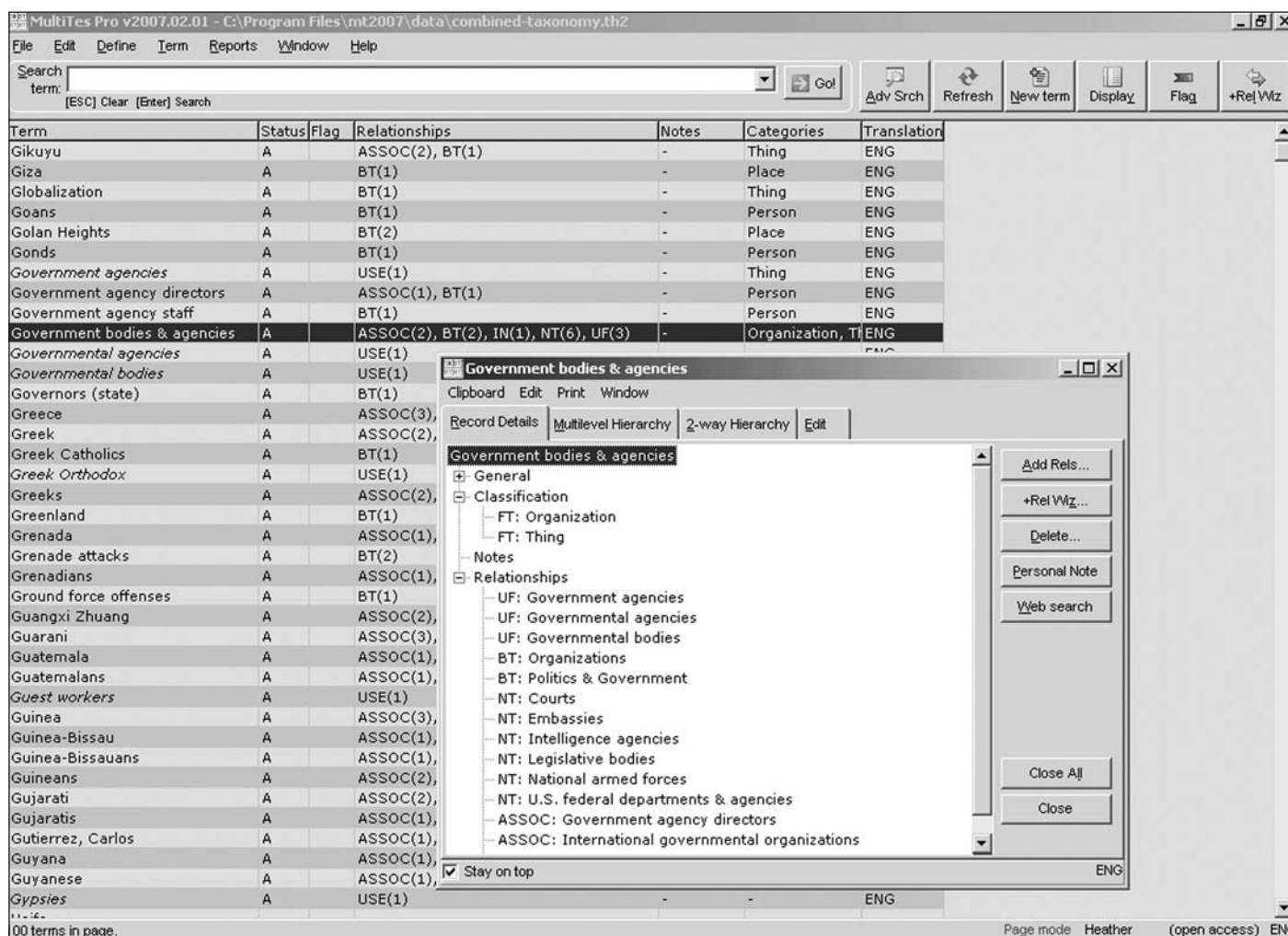


Figure 1 MultiTes Pro screenshot

on the desired term and calling up a new pop-up window, rather than using a screen pane/area. The term window provides the term name, classification, relationships, notes, and other attributes. These details grouped by 'General' (approval status, date, language, etc.), 'Classification' (one or more user-defined categories), 'Notes' (scope notes or other notes), and 'Relationships' (used for broader, narrower, and related). The full hierarchy path of the term can be seen in another tab in the term window. Clicking on an associated term (in either the details or hierarchy tab view) calls up a second display/editing window for that term as well. A series of overlapping term windows can result if the user does not individually close them.

In MultiTes, once a term's window is called up (either by double clicking on the term in the thesaurus list or upon creating a new term), the term and its status can be edited by selecting Edit from the term window's menu. Similarly, relationships, categories, and notes can be added via the Add Relationship button. Although the term display logically groups the general attributes, classification, notes, and relationships, when it comes to adding such attributes, notes, categories, and relationships, their labels are combined in a single alphabetical drop-down list of their abbreviations. As some of the abbreviations may begin with

the same letter, selecting the correct relationship abbreviation may involve additional scrolling. Another drawback is that the scrollable window of relationship-type abbreviations is limited and cannot be resized, which necessitates scrolling. Otherwise, the feature is easy to use.

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 thesaurus, a pop-up window alerts the user and asks whether to proceed with creating the new term. This is a nice feature, because often it is difficult for the taxonomist to remember if a term to which a link is to be made already exists or not.

New terms can be created either from an existing term and creating a relationship, as explained above, or from scratch upon selecting Term > New Term (Ctrl+N) from the menu. The main term-editing drawback is that relationships cannot be changed in a single step, but rather have to be deleted and new ones added.

Thesaurus searching

Searching the thesaurus in MultiTes is quite simple. The user types the start of the term, with truncation assumed,

and hits Enter. There is also the option in preferences to change the search to automatic type-ahead scrolling and specify the number of characters to trigger the search.

There is also an Advanced Search (Ctrl+S) option, available either from the Term menu or by a button. From a drop-down menu in the search window, the user can select from terms which have any of the following characteristics: start with 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 term record. Another option is copying the list of terms to the clipboard for offline manipulation.

User-defined relationships and attributes

MultiTes allows unlimited user definition of relationships, as long as they fall into one of the three main categories of hierarchical (broader/narrower or parent/child), associative (related terms), or equivalent (use/use for). The creation of reciprocal terms is also enforced. Making use of the Define menu, after selecting the general relationship type, the user decides what to call the relationships, their reciprocals, and their abbreviation codes.

Similarly MultiTes allows user-defined notes and categories for terms. For the notes field, one can designate, for example, scope notes to aid end-users, indexer notes with information to support the indexers, and maintenance/history notes for the reference of the taxonomists. MultiTes does not support additional user-defined attributes, but there is no reason that the notes field cannot be used for any purpose, such as a physical address for a company term. In my experience, where a thesaurus required levels of relevancy of nonpreferred terms, I made 'Relevancy' a type of note and a numeric value as the note's content for each term.

Rules enforcement

MultiTes does not allow circular references, but it does permit deleting of terms with narrower terms, which would result in 'orphan' terms. It provides an orphan term report option, though, which can be run for the purposes of periodically cleaning up the thesaurus. A nonpreferred term may have more than one 'use' relationship.

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, simple format explained in the online Help; it does not use tags. This is the only format that the program accepts.

MultiTes has the advantage over the other tools of accepting imported data not merely as files but also simply as text pasted from a clipboard into the blank 'Quick Data Entry' box. This is a practical way for the user to batch-load even just a couple of terms and any relationships at once. As much or as little data as desired can be imported. If a

descriptor term already exists in the database, additional relationships, such as a list of nonpreferred terms, can be added this way.

The features of exporting and generating reports are combined in MultiTes, allowing a full range of displays, subsets of terms to include or exclude, term information to include or exclude, in addition to the different output formats. The output formats available are ANSI text, HTML (as a web page), XML, CSV delimited, or an output directly to the printer. Display formats include alphabetical, top term, hierarchical, subject categories, classified, and rotated index. Terms can be included or excluded by date, approval status, flag, orphans, whether postable or nonpreferred, relationships, categories, notes, etc. The user can also specify what term information to include in the output with respect to relationships, notes, and categories. All terms or an alphabetic range can be selected.

Two add-on products to MultiTes facilitate web and enterprise deployment. MultiTes WDK (Web Deployment Kit) consists of a set of tools which enable publishing of a live MultiTes thesaurus on the Internet/intranet, allowing users with any HTML browser to perform dynamic queries over the thesaurus database. MultiTes EDK (Enterprise Deployment Kit) consists of a set of tools which help the taxonomist transfer and publish thesaurus data (.th2) on corporate servers.

Online help and tutorials

MultiTes has an acceptable standard online Help, which includes Search, Index, and Favorites. Additionally, a tutorial is provided both through the online Help and downloadable from the MultiTes web site.

Term Tree

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 (approximately £370 or US\$700)

www.termtree.com.au

Thesaurus display

Term Tree displays a hybrid alphabetical/hierarchical view of the thesaurus. The terms are listed alphabetically, but with a plus sign to indicate those terms that have narrower terms. Clicking on the plus sign (or double clicking the term), expands the hierarchy for each term to include its narrower terms at the next level. (Additional clicks are needed if there are additional lower levels.) Nonpreferred terms are interspersed among the preferred terms and are displayed in a different font style and color, and with an 'np' icon. The display makes use of various graphical icons next to all terms. The ability to expand the hierarchy below any term within the alphabetical list is a definite plus with Term Tree.

Another nice feature is the View menu which allows the user to limit the thesaurus display to just top terms,

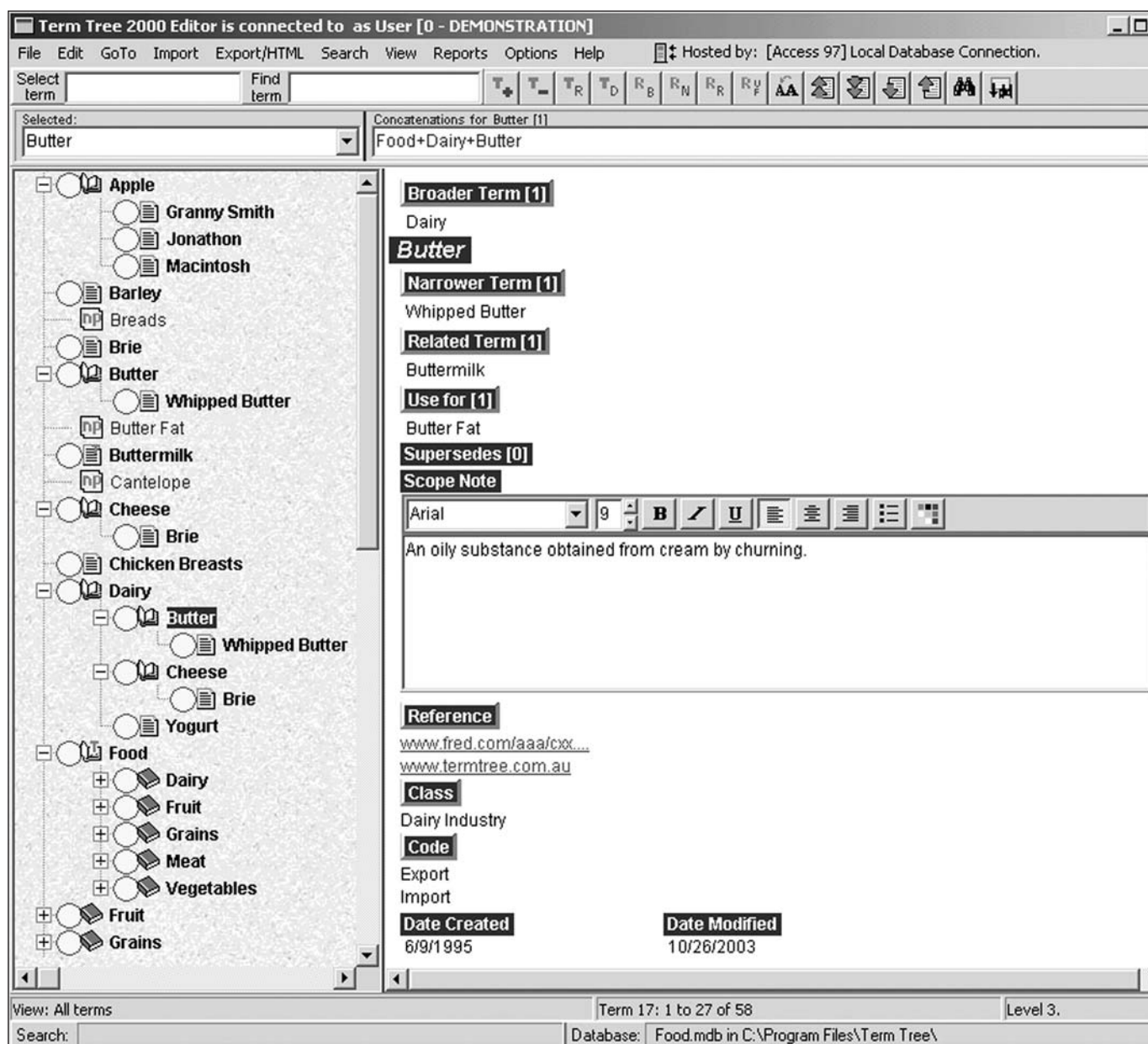


Figure 2 Term Tree screenshot

preferred terms, nonpreferred terms, authorized terms, inactive terms, and/or terms at a specific level. In other products, this can only be done as a report.

Term display and editing

Term Tree supports two different kinds of thesaurus displays, with different kinds of relationships, depending on what type of thesaurus it is. Upon creating a new thesaurus database, the user can specify if it is a subject thesaurus (in accordance with thesaurus standards) or a function/classification thesaurus (in accordance with records management standards). No matter the thesaurus type, Term Tree's main user interface features a split screen with the browsable thesaurus in the left pane and the editable relationships and details of a selected thesaurus term in the right pane.

In Term Tree, the right-hand term detail pane (also known as the relationship pane) includes the following editable fields: term name, its immediate broader term(s), narrower term(s), related terms(s), use for term(s), superseded term (if activated), scope note, reference, class, code, date created, and date modified. Interestingly, the broader term category is displayed in a position above the term name, so as to visually represent the parent aspect of the broader term, while all other relationships and categories are listed below the term. Editing or adding to these relationships requires clicking (left or right mouse button) on the field type label in order to open a pop-up menu.

The scope note field within Term Tree's relationship pane offers full formatting capabilities, including font size, style, face, and color, along with justification and bullets – features that could be considered superfluous for a mere scope note.

The live spell-check in the scope note field is a definite benefit, though.

Term Tree has the advantage over the other tools of permitting the greatest degree of term editing that can be done directly from the left-hand thesaurus browse pane, through keyboard shortcuts (function keys, delete key, insert key, etc.). When a term is selected from the list, keyboard shortcuts allow the taxonomist to rename the term, activate/deactivate it, add a broader term, narrower term, related term, and a nonpreferred ('use for') term, or delete it. A right-click context menu on a selected term within the left-hand thesaurus browse pane also enables some of these functions.

Term Tree provides two methods for creating term relationships: by selecting a relationship-type label within a given term's node details, or by dragging and dropping a term name. In the first method, when clicking upon a relationship-type label for a term, a context menu pops up presenting the options: 1) insert a new term and link it, 2) show a list of existing terms available for linking to, and 3) show a list of terms for possible linkage disconnection. The list of terms available to link to comprises the overwhelming majority of the preferred terms in the thesaurus. For a narrower term link, for example, the terms displayed as available will be preferred terms, not the term itself, and not its broader terms or broader terms of its broader terms. For a broader term link, the terms displayed will be preferred terms, not the term itself, and not its narrower terms or narrower terms of its narrower terms. In this way 'circular' relationships are systematically prohibited. Multiple terms may be checked off at once for simultaneous linking.

The other method of creating relationships, that of dragging and dropping, is achieved by clicking on a large dot next to a term in the thesaurus list so that a term can be dragged from the term list pane to the relationships pane displaying different term. This method, of course, is only for creating relationships between existing terms (unlike using the relationship label's context menu).

New terms can be created by starting from an existing term and creating a relationship, as explained above, or from scratch upon using the keyboard's insert key, calling up a menu with a right click, or by selecting the add term button from the toolbar. Term Tree spell-checks new terms as they are typed in. Dictionaries included are British, American, and Australian English.

An additional relationship feature of Term Tree is 'supersedes,' which is used for keeping track of new terms that replace older terms.

Thesaurus searching

In Term Tree at the top of the screen there are two search boxes: one that searches for the term in the thesaurus list and also brings up the term in the detailed relationship pane, and another search box that will only find and highlight the term in the thesaurus list, but not change the content of the relationship pane. This is so that the taxonomist can drag and drop a term found in the thesaurus to the relationships of a different term in the relationships

pane. In either search box, type-ahead scrolling is supported.

Term Tree additionally permits navigation of the thesaurus from within the menu option of adding a link to an existing term. As mentioned above, a right-click pop-up context menu for broader term, narrower term, related term, and nonpreferred term offers the option of 'Show list of terms available for linked to . . .'. At the top of the list of terms, which could comprise the majority of terms in the thesaurus, there is a search box with type-ahead scrolling search on the start of the term.

Term Tree also provides Search on its menu, which permits searching fields other than term, such as scope note, and allows searches to be limited by source, code, class, and date. If multiple terms match the search, rather than getting a list of matches, as is done in MultiTes, the user needs to select 'next occurrence' from the menu. In my opinion, a list of matches would be more practical.

There is also a find and replace feature, but it is just for the scope notes, which does not seem particularly practical as a search field.

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, etc., creating additional kinds of relationships or types of notes is not supported. The system does, though, 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.

Rules enforcement

Term Tree allows deleting of terms with narrower terms, which would result in 'orphan' terms. But it does offer an orphan term report option, which can be run for the purposes of periodically cleaning up the thesaurus. Circular references are not allowed.

Multiple 'use' relationships are a preset option to permit or not at the time of creating a new thesaurus database. Once selected, though, this permission cannot be changed for the current thesaurus.

Importing, exporting, and reports

Term Tree supports importing data as a tagged ASCII text file, but not XML. Tags, explained in the Help documentation, can define all relationships, date, class, reference, and code. Also through the import menu, Term Tree can import databases from two other thesaurus products, MultiTes and a.k.a, but I was not successful at getting it to work on a MultiTes database with the demo version of Term Tree.

Term Tree can export to the greatest number of file formats: XML, CSV comma delimited, tab delimited, Excel

spreadsheet direct, HTML web site (multiple linked pages), Windows Explorer, netcat (a CSV comma-delimited format for netcat CRM systems), MetaBrowser search vocabulary, and Term Tree's own tag format. Optional additional Term Tree modules (at A\$300 each) allow exporting to certain records management systems: Trim, Objective, RecFind, and PCDocs/HummingbirdRM. The only output format lacking is plain text.

Term Tree's XML export feature includes the ability to create framework CSS (cascading style sheets) and DTD (document type definition) files. For the web publishing feature, which even includes an FTP client, the license for Term Tree only permits web HTML pages to be processed for the website of the registered owner of the Term Tree product. Preparation of web HTML pages for contract clients may only be done if a commercial license is negotiated with This to That Pty Ltd.

Reports can be generated from Term Tree in more types than most users need: hierarchical, concatenation, keyword AAA (a records management standard), term relationships, category listing, classification report, KWIC, KWOC, and classification out of context. The report output format, however, is limited to MS Word. As with MultiTes, a full range of term inclusion/exclusion is possible: by date, approval status, orphans, postable (preferred) or nonpreferred. The user can also specify what term information to include in the report with respect to relationships, notes, and categories. All terms or an alphabetic range of terms can be selected for a report.

Online help and tutorials

Term Tree has an acceptable-standard online Help, which includes Search, Index, and Favorites. In addition, Term Tree includes a tutorial, called a 'learning guide.'

TCS-10

Product name: TCS-10, version 2.26

Product vendor: Webchoir Inc. (Los Angeles, California, USA)

Price: US\$399 Personal edition (single user); US\$1,495 for 10 users

www.webchoir.com

Thesaurus display

Webchoir's TCS-10 (which stands for Thesaurus Construction System) provides two different kinds of thesaurus displays depending on what type of thesaurus it is. Upon creating a new thesaurus database, the user can specify whether it is a hierarchical type or a 'long list' alphabetical type. Even for the hierarchical type of taxonomy, the user still has the option of toggling between the hierarchy and an alphabetical display (not an option if it is designated long list). The alphabetical list displays a section of the taxonomy beginning with a letter of the alphabet. The hierarchical view lists only the top hierarchical terms, which can be expanded to show narrower

terms by either a click on the term or clicking on the 'expand all' button. The 'expand all' and 'collapse all' buttons are nice features.

TCS-10 is the only tool of the three that does not include display of nonpreferred terms in the thesaurus list, in either the alphabetical or hierarchical display. Since nonpreferred terms are still searchable in TCS-10, it could be argued that it is not necessary to have them browsable as well. Excluding nonpreferred terms from the display does make the term list easier to browse, especially in a thesaurus that has more nonpreferred terms than preferred terms.

TCS-10, which maintains its files as MS Access databases, 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 inter-related. The user can alternate between them using the 'Window' option on the file menu.

Term display and editing

Like Term Tree, TCS-10's main user interface features a split screen with the browsable thesaurus in the left pane and the editable relationships and details of a selected thesaurus term in the right pane. A third section at the bottom of the thesaurus pane allows single-click online searching of a selected term through a chosen web search engine.

TCS-10 differs from the other two tools by making a distinction within its right-hand pane (called the node detail area) between a mere display of term details and an editable window of term details and relationships. When a term is selected, an uneditable display of some of the term's details appears in the node detail area. Either 'Edit Current Descriptor' must be selected from the main menu, or an edit icon button must be clicked in order to make changes or additions. This added step to edit a term's details is a minor inconvenience. Furthermore, one must click 'Close' to leave the editable area and return to the regular view such as to select another term or even access items in the Edit menu, also an inconvenience.

The uneditable term details include: nonpreferred terms, broader terms, related terms, scope note, and approval status. What is missing in this display are the narrower terms. On the other hand, TCS-10, unlike the other two products, displays a full hierarchy path of broader terms up to the top term. From the term display, it is possible to jump by means of hyperlinks to broader terms and related terms, but not to narrower terms (not displayed) or nonpreferred terms (displayed but not hyperlinked). Being permitted to jump to some, but not all relationships is a little frustrating.

TCS-10 supports two choices of data entry style in the right-hand term pane: one page editing (introduced as a new feature in a later version of TCS-10 and now the default) and multiple tabs editing. It is good that the user can check off at the top of the page which relationships and attributes should be included in the editing pane, so that data fields not being used do not appear and clutter up the window. Nevertheless, even with some of the attributes excluded, the pane still requires scrolling. Additionally the style of boxed areas is not as user-friendly as the MultiTes or Term Tree edit window interfaces.

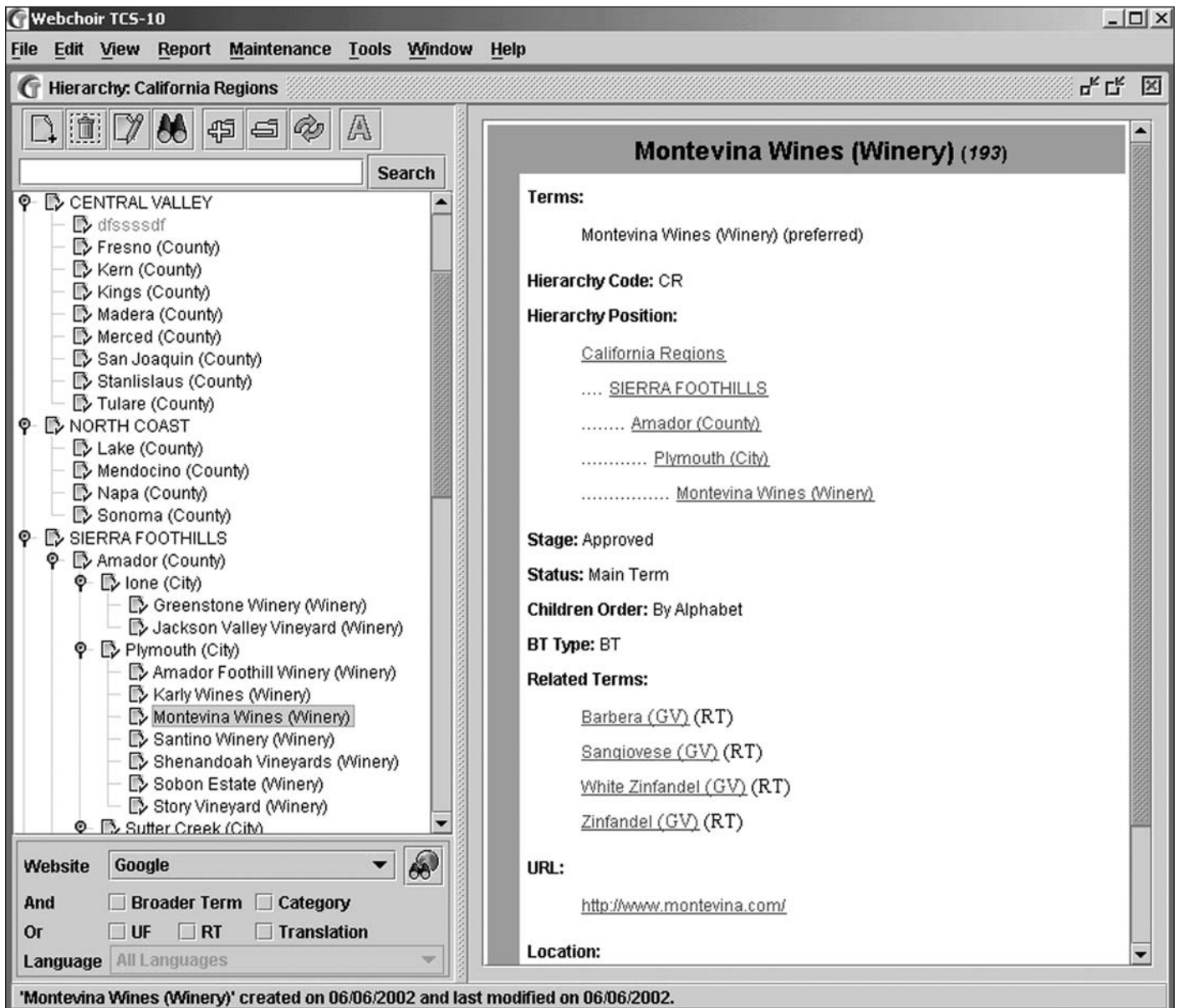


Figure 3 TCS-10 screenshot

In TCS-10, the editable fields for each term include, in order, the following: descriptor info, term name, used for terms, scope note, related terms, used for plus, descriptor images (links to files), term sources, categories, translations, web page URLs, user notes, and user relations. As in Term Tree, a nice feature is the drop-down menu of possible terms to which to link as related terms. While URLs and images are nontraditional bonus attributes, what is blatantly missing here are broader-term and narrower-term relationships. This is apparently because terms can be created only as narrower terms of existing terms, so once this pane appears, a broader term will have already been defined.

Unlike the other products, TCS-10 does not permit multiple standard 'use' references. Instead, it provides a different relationship called 'use for plus' or 'used for and,' which is abbreviated as UFA. The implication is that this should be used to direct the end-user of the thesaurus from a single compound nonpreferred term to two or more

preferred terms. For the use for plus relationship, a drop-down list of existing nonpreferred terms is available to select from.

A benefit of TCS-10, on the other hand, is the option of 'move branch' which will simultaneously remove one broader term and replace it with a different broader term.

Although lacking keyboard shortcuts, TCS-10 also permits editing of terms directly from the left-hand thesaurus browse pane. In fact, TCS-10 has the most extensive right-click context menu for each selected term. The pop-up menu items call up windows that allow term renaming, deleting, and the adding of broader terms, narrower terms, related terms, nonpreferred terms, different kinds of notes, reference sources, and categories. From the pop-up menu there is also the ability to move the branch, edit the user relation, and move the order of narrower terms.

Unlike MultiTes and Term Tree, in TCS-10 there is only one

method of creating a new term. An existing term must first be selected from the right-hand thesaurus pane. Once selected, Insert Term can be selected from the Edit menu. If a term is to have more than one broader term, then Add Additional BT must be selected from the Edit menu. While creating terms first as narrower terms to existing terms is the most common method by which taxonomists work, it should not be the only method permitted. Sometimes a taxonomist might want to create a new term first as a related term to an existing term.

TCS-10 has a limited drag-and-drop feature. Unlike Term Tree, a term cannot be dropped to add any relationship, rather it can only be moved from having one broader term to having a different broader term. This is done through the thesaurus tree display pane. If the destination broader term is not visible in the display, the procedure cannot be done, so in a large hierarchy many broader terms will need to have their hierarchies collapsed first so that the target term will appear in the window display. The practicality of the drag-and-drop feature is therefore somewhat limited.

Thesaurus searching

As with MultiTes and Term Tree, TCS-10 features both a search box at the top of the scrollable thesaurus list and an advanced search feature from the menu. The search box, unlike MultiTes and Term Tree, does not support truncation. On the other hand, text within this search box does not merely match the start of a term, but also can match a word within a term. As this kind of search can yield multiple results, a window pops up with all matching results, a feature found only in the advanced search of MultiTes, from which any matched term can be selected and jumped to.

The advanced search (Ctrl+F) feature from the Edit menu, and also via an icon button, does not have as many limiting options as the advanced search in MultiTes or Term Tree. The searchable fields are only the term name (descriptor) and scope note. However, TCS-10 uniquely has the advantage of offering guided Boolean search. Options include search Within Results (AND), Add to Results (OR), and Not in Results.

User-defined relationships and attributes

TCS-10 supports unrestricted user-defined types of use/used for relationships and types of related-term relationships (through the Maintenance menu's Lookup Table) but restricts the broader/narrower relationships. Instead, it provides the option to choose from three additional types of broader/narrower relationships, all of which are in accordance with ISO standards. They are defined in the system with the abbreviations BTG, BTP and BTI, which can be selected from a drop-down list. BTG stands for broader term – generic, which is when a narrower term that 'is a' member or type of broader term. BTI stands for broader term – instance, when a narrower term is a specific (proper noun) instance of a generic broader term. BTP stands for broader term – partitive, which is when a narrower term is part of a broader term whole. If the user has more specific thesaurus relationship requirements, the limited choice of standard relationship types can still be useful. For example, a geographic hierarchy is actually a kind of whole-part relationship, so the BTP relationship could be

used to relate cities to their states and states to their countries. I consider the BTI relationship particularly valuable, since named entities have their own particular attributes and may need to be set off differently from generic broader/narrower term relationships.

TCS-10 also supports user-defined notes and other attributes. The note types are in addition, not instead of, the system-supplied Scope Note field. Unlimited user-defined attributes can be created for terms. Called ambiguously 'User Relation Type,' these attributes could be utilized, for example, for location or contact information of an organization or person term. A feature unique to TCS-10 is that one can even create authority control for a relation type, such as a lookup table of industry codes or state abbreviations.

Rules enforcement

TCS-10 does not permit the creation of a relationship to a term that does not yet exist. TCS-10 also includes user-defined systems options (from the Edit menu) to allow/disallow duplicate terms and to allow/disallow deletions of terms with narrower terms. The user can also specify whether narrower terms should display under their parent term in an order other than alphabetical.

Importing, exporting, and reports

TCS-10 supports the greatest number of formats for importing data. From the menu, Tools > Import, the choices are ASCII, XML, and MARC. For the import of ASCII text files there are additional submenu options for importing a simple descriptor alphabetical list, a standard thesaurus list with notes and relationships for each term, and a simple hierarchical list using periods for indents. There is also an ASCII import wizard to aid the process. For importing XML, a tag mapping table needs to be completed.

TCS-10 supports exporting as XML, ASCII, MARC, HTML (with hyperlinks for the web), and a proprietary format for Webchoir's other indexing and search products. Additionally, user-created look-up tables can also be exported. HTML export options include web pages with or without frames.

Report options from TCS-10 are comparable with the other thesaurus tools and include alphabetical, hierarchical, parallel hierarchy, hierarchical portion, rotated, hierarchy list, and descriptor by category. Terms can be included/ excluded by date, approval status, or alphabetic range. For the alphabetic report, all relationships, notes, categories, attributes, etc. can be included or excluded in the output. Additional kinds of reports include top term list, used-from (UF) list, URL list, image list, category list, and candidate term list. The reports are output as either ASCII text files or HTML. Unlike the output of TCS-10's export feature, the reports generated in HTML are simple HTML files lacking term hyperlinks.

Online help and tutorials

TCS-10's Help, while comprising a reasonable amount of information, is not easily searchable. The bulk of it is on

merely eight pages, one for each menu item. So, when searching on Help, the user is returned to one of these pages and must scroll through the long page to find the highlighted search term. Furthermore, the online Help file has only Contents and Search, and lacks an index. The Help text also displays too wide for easy reading. In addition to the online Help file, the same Help documentation is also in an HTML file.

Summary

There is no clear-cut best or worst thesaurus software. Each tool fulfills the basic requirements. Different user preferences and different taxonomy projects will determine which product to choose. In building a thesaurus, speed and ease are important. Each tool permits adding successive relationships of the same type (nonpreferred terms, narrower terms, or broader terms) simply by hitting the enter key.

MultiTes is a good-value thesaurus creation tool. Its basic features are easy to use, but some of the reporting and batch methods of making changes are not so intuitive. I especially like the way it is easy to define unlimited types of relationships and is simple to import terms with relationships. Compared with the other tools, MultiTes lends itself most easily to offline text file importing and exporting. Its major drawbacks are the inability to subsume or merge terms in a single step and its lack of a spell-check capability.

Term Tree is graphically not so clean and attractive. I feel

that there are too many unnecessary icons, but the software is very functional. The taxonomist can add relationships to a term selected from the browser in just one step/click, unlike the other two packages. Term Tree serves efficient term entering and editing by making use of keyboard shortcuts and single-action toolbar buttons. Whether the feature of drag and drop is really that important is another matter.

TCS has the advantage of supporting the dual level of thesaurus databases and named hierarchies within those thesaurus database. This is a useful feature for creating multiple facets or top-level terms for a single thesaurus project. Terms can have relationships across different hierarchies, including broader terms in more than one hierarchy. My main complaint about TCS is that the broader/narrower term relationships are not part of the editable relationships of a term and that the entire thesaurus must be created from the top down.

Reference

Hedden, H. (2007) Controlled vocabularies, thesauri, and taxonomies. *Indexer* 26(33) (March), 33–4.

Heather Hedden is a taxonomist with Viziant Corporation, a continuing education instructor with Simmons College Graduate School of Library and Information Science, and the manager of the Taxonomies & Controlled Vocabularies SIG of the American Society for Indexing. E-mail: heather@hedden.net

Some indexer-authors

SI Scottish Group member Margaret Christie had her first volume of poetry published as a chapbook in autumn 2007. (Margaret Christie, **The oboist's bedside book**. Glenrothes: HappenStance, 2007. 36pp. ISBN 978-1-905939-11-4 £4.00) The poems provide a fascinating insight into the production of sound, as it is shaped into music by the artist's technique, endless practice and the sheer joy of playing. Some poems have featured in previous collections, but this book heralds the start of solo volumes.

Margaret summarizes her activities thus: 'Margaret Christie indexes, copy edits and proofreads to support her oboe habit' She even includes references to indexing in her poems!

Another distinguished member of the Scottish Group, Colin Will, is also a published poet.

Hazel Bell, an advocate and practitioner of fiction indexing, has published cumulative indexes to A. S. Byatt's novels (Hazel K. Bell, **The Frederica Indexes: Cumulative Indexes to A.S. Byatt's Novels 'The Virgin in the Garden'/'Still Life' and 'Babel Tower'/'A Whistling Woman'**. HKB Press, 2007. 21 pp. ISBN-13: 978-0955250361.£6.00); and see the advert on this page for another of her publications.

OAK KNOLL PRESS AND HKB PRESS *present*



From Flock Beds to Professionalism

A History of Index-Makers

by Hazel K. Bell

Preface by David Crystal

- Describes the history of index-making since the fifteenth century
- Presents brief biographies of 65 important index-makers
- Outlines the history of indexing groups and societies through 1995

Hardcover with dust jacket
6 x 9 inches, 340 pages



Available in North America
from Oak Knoll Press: ~~\$95~~
Pre-publication special: \$65

1-800-996-2556
orders@oakknoll.com
www.oakknoll.com

Available in the UK from
HKB Press: ~~£49~~
Pre-publication special: £33

+44 1707 265201
www.aidanbell.com/
hkbpress/welcome.htm