

White Paper: Aligning the Transportation Research Thesaurus with Expanded Transportation Business Needs

September 27, 2007

Background

The Transportation Research Thesaurus (TRT) is a controlled vocabulary currently maintained through the Transportation Research Board (TRB) and several contractors/consultants. Improvements to the TRT have been funded through the National Cooperative Highway Research Program projects.

The original intended use of the TRT was to serve as an indexing language for electronic documents housed in the Transportation Research Information System (TRIS). Currently, the TRT is being used to index electronic documents beyond its original purpose including web pages, and electronic documents not housed within TRIS. Recently the TRB Committee on Library and Information Science for Transportation's (LIST, ABG10) Subcommittee on the TRT agreed that the TRT is an appropriate indexing language for use by state DOTs to index web pages and other documents not within the transportation research domain. They also agreed that the TRT should be marketed for these uses.

With this extended use come new challenges and questions for the management of the TRT.

Washington State Department of Transportation (WSDOT) is using the TRT to index research reports and engineering publications in a proof of concept project to improve findability and retrievability within these document collections. WSDOT is also promoting the TRT as a controlled vocabulary to be used in various document management initiatives within the agency. As we have applied the TRT to these reports and to a catalog of agency databases, we have identified some limitations of the current TRT that we would like to bring forward to the managers of the TRT process.

Current Management Process

As the TRT has begun to be used beyond the purpose it was designed for, the methods and processes for term selection specific to the domain have stayed the same. Current term selection methodology follows the standards outlined by the National Information Standards Organization (NISO) in ANSI/NISO Z39.19-2005 *Guidelines for Construction, Format and Management of Monolingual Controlled Vocabulary*. Suggested terms may be contributed to the TRT by anyone through the website. Once a number of terms have been submitted, they are reviewed by a contracted lexicographer and a determination is made whether to include the term into the TRT, including its placement within the structured vocabulary. This determination is made, based in large part but not exclusively, on review of the literature in TRIS. This decision is sent to the LIST TRT subcommittee for review and comment. Decisions are finalized after reviewing this input. This type of review for identifying terms is called a Literary Warrant and has made sense for the TRT as conceived to be an indexing structure for TRIS.

Meeting the Needs of the Industry

Now that uses of the TRT extend beyond indexing TRIS documents, the process for identifying terms for use in the TRT should be reviewed to determine whether it addresses the needs of the broader transportation community. Options to consider in determining new terms are Literary, User and Organizational Warrants. For purposes of common understanding of this document, they are defined as follows:

Literary Warrant is a justification for selecting terms based on a significant frequency of the term in the information resources to be indexed.

User Warrant justifies the selection of terms based on terms that users employ to find information.

Organizational Warrant justifies the selection of terms based on business requirements and the business language used by an organization.

It is clear that use of a Literary Warrant using the literature within TRIS was appropriate for the TRT when it was limited to indexing TRIS. However, TRIS represents a small subset of transportation information resources and the Literary Warrant may no longer be the appropriate choice. Some users, such as WSDOT, are finding that the current Literary Warrant limits the inclusion of terms within the TRT so much that it will not meet our needs as an indexing structure for the types of materials we are trying to index.

Example 1

In a project currently underway at WSDOT to index documents using TRT, some of the preferred terminology seems to have a specific regional context not found in operational and research documents indexed by WSDOT. As an example, for a paper on the subject of the use of parking in relation to public transit, the preferred term available in the TRT based on Literary Warrant is 'fringe parking'. WSDOT queried subject matter experts within the agency and found that the term 'fringe parking' was unused in the region and that the commonly used term, which has both literary and user warrant for WSDOT, is 'Park and Ride'.

While the term 'Park and Ride' is currently in the TRT as a synonymous or lead-in term (called an Equivalent Relationship in the ANSI/NISO standard and designated by USE/UF notation) for "fringe parking". The selection of the preferred term in this example raises a question about how terms are selected for the TRT as a whole.

While both terms are used in literature, this is an instance when the transportation user and organizational communities should have an opportunity to guide the TRT on the best term to use for the transportation communities. As we broaden its use, it is important that the TRT should reflect national terminology and seek to avoid regional biases.

In addition, we also considered the TRT when asked to develop metadata for agency web pages and to establish categories for document management within a developing Project Management Reporting System. The current limited terminology made these applications very limited. These information environments create new demands. What seems to be missing from this format is direct or indirect input from subject matter experts (traffic engineers, design engineers, transportation planners, maintenance technicians, and hydraulics engineers, etc). Adding input from user communities would help ensure that not only the structure of the thesaurus fits the business but also that the terminology is more accurate for the users of the domain. And, it would improve the findability of transportation information. In a recent

presentation at the Library Connectivity Pooled Fund Study Annual Meeting, Peter Young and Gary McCone of the National Agricultural Library (NAL) discussed the National Agricultural Library Thesaurus (NALT). The NAL has seven indexers on staff that provides support to the continued development of the NALT by identifying new terms or other changes in the language in the field of agriculture. They use the subject area experts within the NAL to determine preferred terms and standard usage, and even solicit input and feedback on NALT's structure from international partners in agricultural research. With the growth and expanded use of the TRT, it is time to consider a more sustainable model for continual development of the TRT. A part of this should include the use of subject area experts from across the nation to advise on common usage of terms.

Example

In developing terminology to describe, design, and catalog our information systems in a Data Catalog, WSDOT uses a consensus driven format for terminology development. Subject area experts from the affected business areas have direct input and in many cases veto power over the naming of concepts. This ensures that not only the structure of the thesaurus fits the business but also the terminology is more accurate for the users of the domain.

In this case, the TRT is used as reference for the Business Topics within the Data Catalog but has not had much applicability beyond that.

Gaps in Terminology

In reviewing the TRT, there is a lack of breadth in equivalent relationships. Many preferred terms within the TRT have no equivalent relationships and most others have one or two at most. A thesaurus with more equivalent relationships would allow for a smooth crosswalk between regional vernacular, industry terminology (if different than preferred), and the preferred term within the TRT, allowing for better findability in information retrieval systems.

The indexing project mentioned above also found that the TRT is lacking in breadth of terms for transportation operations. It is important that this facet be built out as we expand beyond the indexing of research related documents.

Managing Scope

While there are terminology gaps in coverage of the transportation field there are also several domains in addition to transportation that exist within the thesaurus. While this may be convenient for indexers to have terms co-located in one thesaurus, it adds a problem for maintenance and introduces potential conflicts with common usage in the user community for that field. For terms related to psychology, mathematics and environment, there are other more authoritative sources for the terminology. We recommend that these sources be used rather than developing terms within the TRT. Using these terms from thesauri maintained within the user domain will maintain consistency with the field, reduce conflicting uses of terms, and reduce TRT maintenance issues.

Managing an Evolving Vocabulary

Use of a controlled vocabulary within the transportation community is important. It will help us link resources between organizations throughout the nation and the world. It has the potential

to help us improve enterprise management of information within our organizations. The key user benefit is more timely and accurate retrieval of information needed for the development, operation, management and improvement of transportation systems. To do this, we need a controlled vocabulary that is consistent throughout the industry, versatile in meeting needs, and agile enough to evolve with the transportation community. The TRT is the foundation of this, but, as currently resourced, unable to develop as rapidly or as broadly as needed. To really achieve this vision, a more stable basis for funding and staffing the TRT is needed.

State of Practice

A quick look at other industries shows a stronger investment in development and management of controlled vocabularies. The National Agricultural Library maintains the [NAL Agricultural Thesaurus](#) (NALT). They have seven staff dedicated to indexing and continual development of the thesaurus. The NALT also draws upon the opinion of their user communities to determine preferred terms. Their thesaurus currently has over 68,500 terms that are used to index over 4 million documents in Agricola.

The National Library of Medicine's controlled vocabulary thesaurus is called [MeSH](#) (Medical Subject Headings). This thesaurus is continually maintained by 11 staff. It currently includes nearly 23,000 terms and over 150,000 supplementary records with thousands of cross references. MeSH is used to index over 4800 professional journals used within the medical community.

Currently, the TRT is managed primarily as a collateral duty with part time paid and part time voluntary professional assistance. This is clearly not a sustainable model nor is it a model conducive to timely evolution of the TRT to adequately address evolving language and user needs.

What's in a Name?

The TRT was originally created to index research documents but the expanded use of the TRT covers much more. The name implies that the TRT is appropriate for a narrow subset of transportation literature. This makes it easy for potential users to dismiss as irrelevant to their work. Given that the TRT is not well known beyond the library community, this is a good opportunity to determine if the name should be changed to reflect the broader application of this controlled vocabulary. Some names to consider are: National Transportation Library Thesaurus (NTLT); or Transportation Information Thesaurus (TINT) or Transportation Thesaurus (TnT).

An Eye to the Future

The model of an indexer cataloging every information resource within transportation quickly becomes infeasible when faced with the tremendous amount of information being created, published and disseminated. In the near term, to help practitioners deal with the ever-increasing volume of information, work to make user interfaces easier to use is critical for implementation of the TRT throughout the transportation industry. This might include more relational terms operating behind the scenes for most users. Or perhaps term selection tools that direct users to appropriate choices through drop down menus.

Library catalogs, a traditional interface between users and information, have evolved from hand-typed cards in drawers to online public access catalogs (OPACS) which employ information science techniques. Behind the scenes, structured subject headings guide users from broader

to narrower, or unused to preferred terms. Automated library systems now augment traditional organizational schemes, such as Library of Congress subject headings, with relevant specialized thesauri. The TRT is currently downloadable into such systems. However, a marketing strategy is needed so that catalog managers in the transportation community are actually taking advantage of this. In addition, investigation should be made into how these information science techniques and thesauri could be used to improve web searching tools and document management systems

There is a need for a nimble protocol that can be adapted to processes within an organization and applied to many uses while retaining adequate structure. . We encourage managers of the TRT to keep an eye on emerging technologies in order to continue to adapt this controlled vocabulary for future use.. Building upon the strong framework already evident in the TRT will allow this valuable resource to have a broader impact throughout the transportation field, and consequently, greatly assist in efforts to capture, organize and distribute the specialized information utilized within the professional community

WSDOT Requests of the NCHRP 20-79 Panel

1. Develop an advisory group of subject area specialists to advise on terminology in the TRT. More than one specialist should be used for each subject and represent different geographic regions.
2. Identify gaps in transportation terminology within the TRT and determine how best to rapidly address those gaps.
3. Review the TRT for the inclusion of terminology that could/should be addressed through thesauri managed by other domains.
4. Review other thesauri and emergent technologies to determine whether new methods of management are imminent and worth investing in.
5. Develop recommendations for a staff and budget that would adequately maintain the TRT and allow evolution to keep pace with changing terminology and uses.
6. Consider renaming the TRT to reflect the broader applications we are striving for.