

Untangling Regulatory Text: Multidimensional Separation of Concerns and Task-Oriented Linking

Jeffrey A. Lasky
+1 (585) 475-2284
jal@it.rit.edu

Jeffrey Sonstein
+1 (585) 475-7315
jeffs@it.rit.edu

Rochester Institute of Technology
102 Lomb Memorial Drive
Rochester NY 14623

ABSTRACT

Regulatory text is a complex network of information distinguished by technical language, large volume, and tangled concepts formed by embedded cross-references. Scattering and tangling add to the cognitive pressures on users of regulatory text. Multidimensional Separation of Concerns (MDSOC) is a software engineering method aimed at untangling source code objects that contain cross-cutting, overlapping, or interacting concerns. The MDSOC approach is to refactor the source code into separated and untangled unidimensional concerns. The result is more understandable source code. This paper explores the application of MDSOC model to regulatory text. In particular, we sketch out how task-oriented linking can be used to untangle large bodies of regulatory text.

Categories and Subject Descriptors

H.5.4 [Information Interfaces and Presentation]: Hypertext/Hypermedia; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval.

General Terms

Design, Human Factors, Theory.

Keywords

Regulatory text, concern space, scattering, tangling, hierarchies, networks, links, small graphs.

1. INTRODUCTION

Compliance to regulation and law often requires deep understanding of how narrowly-focused and complex concepts are applied to specific sets of conditions. Proper interpretation and application of these concepts is sensitive to small changes in circumstances, given the seemingly unending stream of legislative changes, regulatory rulings, judicial opinions, and interpretations. The end result is an expanding space of interrelated regulatory information.

Problem solving in this regulatory information space (RIS) is often a matter of resolving one or more concerns. Typically, information in a RIS is distributed, or *scattered*, and grouped, or *tangled* with other possibly disparate information. Scattering and tangling place additional cognitive pressures on users. As the volume of the RIS increases, the likelihood and extent of scattering and tangling increases as well. The central intellectual tasks at hand are how best to unscatter and untangle regulatory text and how to organize a dynamic RIS for users engaged in task-oriented activities. In our approach, we propose to assist the user to identify, create, and focus on the tiny sub-space, or small graph, containing the needed problem solving information.

2. STRUCTURAL TYRRANY

Information structure creates what Turkle called a master narrative [8] or a singular view. This singular view creates a tyranny of the dominant concern [7]. In the present context, the result is a constrained mental model of information relationships and reduced ability to create task-specific information structures.

In contrast, users need to actively construct new narratives and to follow different relational paths than those provided by author(s). The relations between units of information and the meanings associated with their relational constructs are not stable and are not universal. The momentary role of the user is what determines cognitive groupings and relational mappings for that user. This point of view is not new; as noted in [3], both Vannevar Bush [2] and Ted Nelson [4] have argued and stressed this point in more general terms.

A tree-like table of contents is used universally to display the manifest structure of regulatory text. As the body of text evolves, the latent structures needed by users for comprehension and application could be provided by hyperlinked referential entries. In so doing, however, the initial and neat hierarchical taxonomic structures are overlaid with messy sets of linkages reflecting the presumed views, categorizations, and multidimensional concerns of users. The result of grafting latent referential structures onto manifest hierarchical structures is increased cognitive pressures.

In the end, the choice of a hierarchical or network structure for associative information organization remains an open issue [3]. At least in the regulatory domain, we believe network organization is an attractive compliment or alternative to hierarchical organization. First, users need to untangle cross-cutting or interacting concerns to isolate task-specific information. These concerns are inherently network-like.

Second, linking new information into the task-specific space will be more natural in a network model. Third, dynamic network structures will be a good match for ongoing problem-solving.

3. SEPARATION OF CONCERNS, SCATTERING, AND TANGLING

Separation of concerns (SOC) is a core software engineering design principle initially developed to address the question of how best to decompose software into flexible and maintainable modules [5]. Now three decades later, SOC has motivated the emergence of a new area of software engineering research collectively named Aspect-Oriented Software Development (AOSD). AOSD methods attempt to factor out scattered and tangled concerns. Multidimensional Separation of Concerns (MDSOC), a major subdomain of AOSD, deals with the particularly nasty problems associated with refactoring software containing overlapping concerns which lie along multiple concern dimensions [7]. The impetus for development of the above techniques is to provide assistance to those who must understand and maintain software systems. From the perspective of cognitive pressures due to scattering and tangling, we see a strong parallel between readers of software and users of regulatory text.

Regulatory text presented in traditional, static, single-view information structures are often seen by users as tangled webs of unrelated concepts. By this we mean that from a task-oriented perspective, multiple seemingly unrelated or weakly-related concepts end up being physically grouped together (*tangled*) in the same unit of regulatory text. By the same token, users often need to “look all over the place” in the regulatory text for materials which seem, from their task-oriented perspective, to be related and yet are not grouped together (*scattered*).

When regulatory text is recast as an n-dimensional information space, AOSD motivations, models, and methods are useful in framing the challenge of untangling regulatory text. In particular, we note the potential value of concern space modeling and hyper-multigraphs [6], and hyperspaces [7].

4. TASK-ORIENTED LINKING

Users of regulatory text need to understand concepts and relationships to get their work done. What is needed is a way to allow a user to separate concerns and then to create a specialized task-oriented information space. We suggest that two computing models, MDSOC and relational database theory, are potentially useful here. Through refactoring regulatory text into separated and untangled one dimensional concerns, the latent structure required by the user’s role can emerge, cognitive pressures can be reduced, and understanding can be increased.

In our view, this requires a tiny subgraph generated by the user that focuses on one set of concerns. Initially, the subgraph will be generated with a concern radius equal to one, where concern radius is the maximal degree of separation from the node of central concern. Given the node of central concern, graph theoretic algorithms will be used to generate the subgraph. Due to the need to untangle, unscatter, and separate concerns from the point of view of the task-oriented user-of-the-moment, this subgraph will not be a fixed structure. Relational database theory suggests to us that multiple views can be

obtained by using a subgraph tool that includes the following sparse set of interactive linking operators:

- add(edge)** (creates new relationship)
- delete(edge)** (untangles existing relationships)
- add(node, edge)** (fuse concerns)
- delete(node)** (separate concerns)
- spread(n)** (spreading activation) [1]

5. INITIAL EFFORTS

We have selected the U.S. federal income tax laws as a first regulatory domain. The U.S. Tax Code is anecdotally recognized to be the most complex body of extant regulatory and legal text. The Code is approximately 15,000 pages in printed form, and contains in graph form thousands of nodes and tens of thousands of edges. The user community includes thousands of government and private attorneys, government tax administrators, and professional tax practitioners.

The Tax Code is subject to frequent revisions and reinterpretations from sources such as the U.S. Congress, IRS rulings, and judicial decisions. In sum, the scale and volatility of the Tax Code, and the need for a multiplicity of views, suggests that the Tax Code is an ideal initial research vehicle.

As a first step, we are analyzing the structure of the Code to generate visualizations of the Code’s scattering and tangling structures. We anticipate that these visualizations will provide insights into the deep structure of regulatory text, which in turn will lead to categorizations, operational constructs, and measurable properties.

6. REFERENCES

- [1] Anderson, J.R. and P.L. Pirolli, Spread of activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10 (1984), 791-798.
- [2] Bush, V., As we may think. *The Atlantic Monthly*, (176, 1), July 1945, 101-108.
- [3] Erin, N. and K. McCurley, Untangling Compound Documents on the Web. *Proceedings of HT '03* (Nottingham UK, 2003), ACM Press.
- [4] Nelson, T.H., Embedded markup considered harmful. <http://www.xml.com/pub/a/w3j/s3.nelson>.
- [5] Parnas, D.L., On the Criteria To Be Used in Decomposing Systems into Modules, *Commun. ACM*, 15 (1972).
- [6] Sutton Jr., S.M, and I. Rouvellou, Issues in the Design and Implementation of a Concern-Space Modeling Schema. *Workshop on Advanced Separation of Concerns in Software Engineering. Proceedings of the 23rd International Conference on Software Engineering (ICSE)*, (Toronto, Canada), 2001.
- [7] Tarr, P., H. Ossher, W. Harrison, and S.M. Hutton, N Degrees of Separation: Multidimensional Separation of Concerns. *Proceedings of the 21st International Conference on Software Engineering (ICSE)* (New York NY, 1999), ACM Press, 107-119.
- [8] Turkle, S., *Life on the Screen: Identity in the Age of the Internet*. Simon & Schuster, New York NY, 1995