

Regulatory Information Management and Compliance Assistance

Shawn Kerrigan, Charles Heenan, Haoyi Wang, Kincho H. Law and Gio Wiederhold
Stanford University
Stanford, CA 94305-4020
Email Contact: law@stanford.edu

1 Abstract

The REGNET Project aims to develop a formal information infrastructure for regulatory information management and compliance assistance. This paper discusses three components of current research and development efforts. The first is a document repository containing federal and state regulations and supplemental documents. This repository includes a suite of concept hierarchies that enable users to browse documents according to the terms they contain. The second is an XML framework for representing regulations and associated metadata. The XML framework enables the augmentation of regulation text with tools and information that will help users understand and comply with the regulation. The third component is the creation of a compliance assistance system built upon the XML framework. The compliance assistance system and the document repository can serve as a backend for the development of application-specific compliance guidance systems. The prototype effort for the document repository has been focused on environmental regulations and related documents. The compliance assistance system is illustrated in the domain of used oil management.

2 Background

Industrial production activities that produce byproducts classified as hazardous waste must comply with both the US federal and state regulations regarding the handling and fate of such materials. Both federal and state EPAs, as well as local governments, impose strict regulations on the treatment and disposal of such chemical wastes. These environmental regulations are complex and voluminous. The regulations can be disproportionately burdensome on small businesses, since these businesses often do not have the resources to staff personnel trained to deal with these complicated regulations and procedures [1, 2]. Many government regulations are now available online. However, most of current online portals are primarily designed for displaying the information for experienced users and are difficult to use for further processing. Information technology (IT), if properly designed and developed, has the potential to mitigate and help solve many of these challenges. Through the application of advanced information technologies and the development of new methodologies, the REGNET project aims to develop a formal information infrastructure for regulatory information management and compliance assistance.

3 Document Repository

One of the objectives for the REGNET information management infrastructure is the development of a document repository for environmental regulations. The scope of our current prototype development covers Code of Federal Regulations Title 40 (40 CFR): Protection of the Environment, along with selected supplementary and supportive documents that focus on regulations covering hazardous waste and the management of used oil. Supplemental documents are important because they often contain information that is necessary for the accurate interpretation of the federal regulation(s) to which they refer [3]¹. Supplemental documents may come in the form of administrative decisions, guidance documents,

¹ Supplementary and supportive documents are important part of regulatory information. To illustrate, for the case “Beazer East, Inc. v. U.S. EPA, Region III”, Beazer East, Inc., argued that the aeration basins were “tanks”, not “surface impoundments”, and were therefore not subject to RCRA’s groundwater-monitoring regulations. The court ruled in favor of EPA by considering the so called “Weddle memorandum”, which was issued to clarify the definitions of a “tank” and a “surface impoundment,” as an interpretive rule, which can be exempted from the notice and comment requirements of the Administrative Procedures Act.

court cases, letters from the general counsel and letters of interpretation from the EPA. The REGNET document repository is designed to make these important documents more accessible. The contents of the repository are available through the mediation of one or more searchable concept hierarchies, or through a regulation assistance system described later in this paper.

4 XML Regulation Framework and Metadata

We have developed an XML framework for environmental regulations. The framework is document centric and includes XML tags for each level of regulation text – for example part, subpart, section or subsection – that mirrors the standard structure of regulations. Parsing systems have been built to transform federal and state environmental regulations from Portable Document Format (PDF) and HTML into the REGNET XML framework.² With XML, it is possible to augment a regulation with various types of annotation and regulation-specific metadata rather than simply to structure the regulation according to how it should be displayed. With respect to the document repository, the metadata types currently added to the regulation framework include concept tags, reference tags and definition tags (see Figure 2).

The concept tags allow the dynamic generation of links to related supporting documents in the document repository. This is useful because supporting documents and regulations may not directly reference each other even when they address the same topic. The automatic application of concept tags to the XML framework means that as new supporting documents are added to the document repository, regulations stored in the framework can automatically be linked to them via the terms that they share in common. Concept tags can be generated “semi-automatically” using existing text mining and information retrieval tools [4]. Currently, we use software from Semio Corp. to extract, clean and define over 65,000 concepts for the 40 CFR regulations and to categorize the concepts according to different interests and applications.

Regulation provisions tend to contain a large number of casual English references to other provisions. These references are cumbersome to look up manually, and they reduce the readability of the regulation text itself. Simple references (for example, “as stated in 40 CFR section 262.14(a)(2)”) and complex references (for example, “the requirements in subparts G through I of this part”) exist throughout the regulations. Given the large volume of federal and state environmental regulations, a manual translation of references would be too time consuming. A parsing system was developed using a context-free grammar and a semantic representation/interpretation system that is capable of tagging regulation provisions with the list of references they contain. Instead of building hyperlinks, which tie the reference to a particular source for the referred document, the reference tags provide a complete specification for *what* regulation provision is referenced. *Where* the regulation is located is not specified so that a viewing system may select from any document repository of regulations to retrieve the referenced provision. This gives better flexibility than a rigid hyperlink structure for maintenance and scalability.

The large number of domain-specific terms and acronyms that appear in regulations can make regulation text difficult for novices to understand. Definition tags allow a regulation viewing system to incorporate explicit definitions of terms and acronyms into its user interface. Presently, the definitions are extracted from the regulations and attached to the terms identified by a parser.

5 Compliance Assistance System Infrastructure

There has been a push by the executive office that government agencies put more emphasis on compliance assistance in lieu of enforcement to encourage companies to comply with regulations [5, 6]. Specialized modules, using expert system technologies, have been built for specific applications and

² The HTML regulations are downloaded from the National Archives and Records Administration’s e-cfr website at <http://www.access.gpo.gov/ecfr/>

business types [7]. In these systems, references to the regulations are not explicitly linked. Our research on developing a compliance assistance infrastructure builds upon the XML regulation framework and takes advantage of the regulation metadata described earlier.

Besides the concept, reference and definition tags, we add logic and control processing metadata to the REGNET regulation framework. Logic metadata comes in two variations. There is only one form of control processing metadata. Regulation logic metadata represents a rule or concept from a regulation using First Order Predicate Calculus (FOPC) logic sentences. These logic sentences are used to represent the rules that must be followed for an entity to be in compliance with the regulations. User interface logic metadata uses FOPC logic sentences to represent compliance questions and a list of possible user answers to those questions. Control processing metadata provides information about what provisions of a regulation need to be checked for compliance. Each type of logic or control processing metadata can be associated with any regulation provision in the document. In the REGNET framework, these three types of metadata are necessary for the system to be able to verify compliance with a regulation. However, they must be specified by a domain expert as they cannot be generated automatically. For the purposes of demonstration, a used oil regulation (40 CFR 279) has been manually tagged with regulation logic metadata, with user-interface logic metadata, and with control processing metadata.

We built a regulation assistance system (RAS) to demonstrate how the regulation meta-data can be used. The RAS functionality is implemented by a web interface that communicates with a compliance checking system. The compliance checking system interacts with a theorem prover component. The structure of this system is shown in Figure 1. The compliance checking system controls the process used to check for violations. First, it parses the XML-structured regulation to extract the information necessary to run a

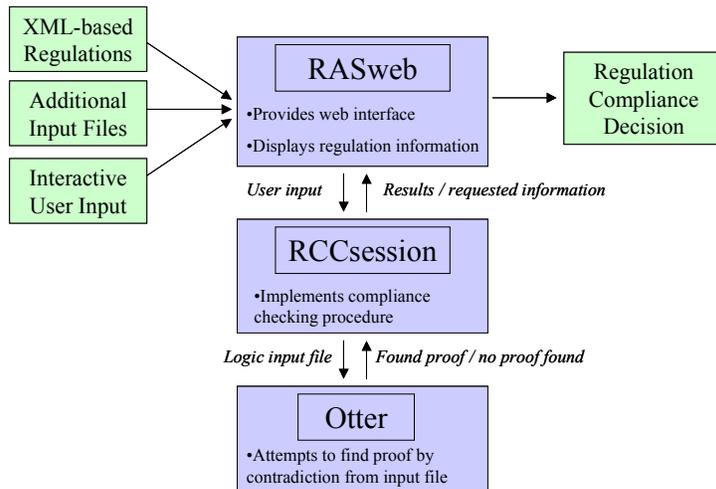


Figure 1. Diagram of the Regulation Assistance System's structure

compliance check. The XML structure allows the system properly to scope the meta-data and to reduce the amount of extraneous data passed to the reasoning system. Only the logic and control processing metadata necessary for the compliance check are acquired and dynamically loaded into the reasoning system. This is important because the performance of FOPC theorem provers decreases rapidly as the number of logic sentences used for reasoning increases. The system design is such that any FOPC theorem prover can be used to perform the logic checks. Presently, we employ Otter, a publicly available theorem prover developed at the Argonne National Laboratory [8].

One essential feature of this web-based compliance assistance system is that it helps guide the user through the regulations. In order to facilitate greater understanding of the regulations, the system makes available a number of enhancements while guiding the user through a compliance check, utilizing the metadata tagged with the regulations. The system can automatically insert links to any referenced regulation provisions and display terms and definitions. Key conceptual phrases for the provision are displayed and linked, enabling instant access to repository documents related to the provision. Options for exploring different scenarios are offered by allowing users to fork the compliance process along all possible paths at any time. When the system completes a check against the regulation provisions or

detects a conflict between the user's answers and the regulation, it displays a summary of the question-and-answer history as well as the results of the compliance check. The use of concepts, definitions, and references is shown in Figure 2. Downloadable logs of completed compliance checks allow users to maintain detailed records of their compliance checks, a feature that should be of value to companies when revisiting the regulations at a later date. The logs of compliance checks can also be uploaded and edited for future compliance checks against the same or updated regulations.

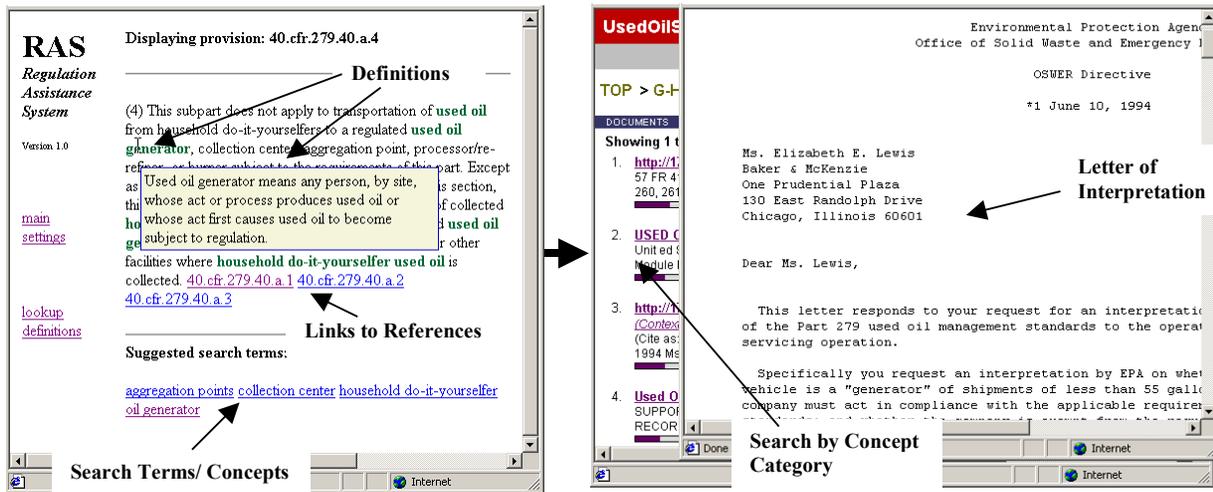


Figure 2. Definition, reference and concept usage

The compliance problem from the perspective of the regulated community can be broken-down into two parts. First, one must determine the set of regulations with which one must comply. Second, one must determine what needs to be done to comply with those regulations. The RAS system primarily addresses the second of these two steps by guiding users through regulations. The RAS system was designed, however, such that it could be used as a component in a larger system that would first assist a user in identifying those regulations that need to be investigated. The RAS system can initiate compliance checks at any point within a regulation, and a compliance check can be started by connecting to the RAS system with a target regulation in the URL. To demonstrate how one could build a compliance guide for a specific application utilizing the RAS system and the document repository as a back end, a sample online guide was built for vehicle maintenance shops. The vehicle maintenance shop online guide was adapted from a paper-based guide developed by the New York State Department of Environmental Conservation Pollution Prevention Unit [9]. Figure 3 illustrates how the demonstration system links into the RAS system to make use of the used oil regulations. The guide targets vehicle maintenance shops and explains what regulations apply to typical work done in that industry. While the original paper-based guide explains requirements and references applicable regulations, our online adaptation provides the additional feature of enabling users to click on referenced regulations and check for compliance by stepping through the regulation itself. Online regulation guides such as the vehicles maintenance shop example located anywhere on the Internet can build upon the compliance-checking capabilities of the RAS system simply by passing target regulations in the URL.



Figure 3. Linking industry-specific guides to the regulation assistance system

6 Related Work

Representation of laws and regulations has been an active research area for decades. There has been a great deal of work on building expert systems for the law [10, 11]. T. Bench-Capon provides a review on the applications of knowledge-based systems for legal applications, particularly the research and development efforts related to the Alvey DHSS Demonstrator project in U.K. [12]. The reference includes several hundred citations that appeared before 1990 that are related to logic and rule based approaches and their application in legal systems. Much of the earlier work in IT and law focused on building systems to optimize decisions with respect to laws, particularly tax law [13]. The current state of legal informatics (i.e. information technology in law) has been discussed by Erdelez and O'Hare [14]. Some of the recent work has focused on investigations into case-based reasoning and information retrieval [15, 16]. Methodologies on tailoring legal documents to users' needs have also been studied [17]. While legal knowledge representation and reasoning has been an active research topic [18, 19], an integrated approach covering the management of regulations, efficient access and retrieval of documents and tools for compliance checking is missing. This research investigates the issues related to the development of a formal regulatory information management system that can also support compliance assistance.

7 Summary

The goal of the REGNET Project is to develop an information infrastructure for regulatory information management and compliance assistance. This paper describes some of the work that has been done to date on creating a regulation document repository, on developing an XML-based framework for representing regulations and associated metadata, and on creating a compliance assistance system built upon the REGNET XML framework.

8 Acknowledgements

This research project is sponsored by the National Science Foundation, Contract Numbers EIA-9983368 and EIA-0085998. The authors would like to acknowledge a "Technology for Education 2000" equipment grant from Intel Corporation and the support of Semio Corporation in providing the software for this research. The authors would also like to thank Professors Barton Thompson and Jim Leckie for their valuable suggestions in this project.

9 References

1. Rechtschaffen, C., "Competing Visions: EPA And The States Battle For The Future Of Environmental Enforcement," *Environmental Law Reporter*, 2000.

2. Romine, M., "Politics, The Environment, And Regulatory Reform At The Environmental Protection Agency," *Environmental Lawyer*, 1999.
3. Heffron, F.A. and N. McFeeley, *The Administrative Regulatory Process*, Longman, 1983.
4. Heenan, C., *Manual and Technology-Based Approaches to Using Classification for the Facilitation of Access to Unstructured Text* (Unpublished Manuscript), Engineering Informatics Group, Stanford University, January, 2002. (available at <http://eil.stanford.edu/regnet>).
5. *National Compliance Assistance Providers Forum*. co-sponsored by the U.S. Environmental Protection Agency and Texas Commission on Environmental Quality: San Antonio, TX., Dec., 2002.
6. *Business Compliance One Stop Workshop*, Small Business Administration, Queenstown, MD, July 24-26th, 2002.
7. Botkin, A., "Wizards, Advisors and Websites, Oh My! Interactive Electronic Tools for Compliance Assistance," presented at the *National Compliance Assistance Providers Forum*, co-sponsored by U.S. Environmental Protection Agency and Texas Commission on Environmental Quality, San Antonio, December, 2002.
8. McCune, W.W., *Otter 3.0 Reference Manual and Guide*. ANL-94/6, Mathematics and Computer Science Division, Argonne National Laboratory, 1994.
9. *Environmental Compliance and Pollution Prevention Guide for Vehicle Maintenance Shops*, New York State Department of Environmental Conservation Pollution Prevention Unit, 2002.
10. Wahlgren, P., *Automation of Legal Reasoning*, Kluwer Law and Taxation Publishers, 1992.
11. Zeleznikow, J. and D. Hunter, *Building Intelligent Legal Information Systems: Representation and Reasoning in Law*, Kluwer Law and Taxation Publishers, 1994.
12. Bench-Capon, T.J.M., *Knowledge Based Systems and Legal Applications*. The APIC Series 36, Academic Press, 1991.
13. McCarty, T., "Reflections on Taxman: An Experiment in Artificial Intelligence and Legal Reasoning," *Harvard Law Review*, 1977.
14. Erdelez, S. and S. O'Hare, "Legal Informatics: Application of Information Technology in Law," in *Annual Review of Information Science and Technology*, M. E. Williams(ed.), ASIS, Vol. 32, 1997.
15. Stranieri, A. and J. Zeleznikow. "The Evaluation of Legal Knowledge Based Systems," *Proceedings of the Seventh International Conference on Artificial Intelligence and Law*, pp. 18-24, 1999.
16. Brüninghaus, S. and K.D. Ashley, "Finding factors: learning to classify case opinions under abstract fact categories," *Sixth International Conference on Artificial Intelligence and Law*, Melbourne, Australia, ACM Press, 1997.
17. Royles, C.A. and T.J.M. Bench-Capon, "Dynamic Tailoring of Law Related Documents to User Needs," *9th International Workshop on Database and Expert System Applications*, IEEE, 1998.
18. *Proceedings of the 7th International Conference on Artificial Intelligence and Law*. Oslo, Norway, ACM Press, 1999.
19. *Proceedings of the 8th International Conference on Artificial Intelligence and Law*. St. Louis, Missouri, ACM Press, 2000.