A PROTOTYPE STUDY ON ELECTRONIC-RULEMAKING

Gloria T. Lau and Kincho H. Law

Department of Civil and Environmental Engineering

Stanford University, Stanford, CA 94305, USA

I N T R O D U C T I O N

The making of government regulations represents an important communication between the government and citizens. During the process of rulemaking, government agencies are required to inform and to invite the public to review the proposed rules. Interested and affected citizens participate by submitting comments accordingly. Electronic-rulemaking, or e-rulemaking in short, redefines this process of rule drafting and commenting to effectively involve the public in the making of regulations. The goal of the e-rulemaking initiative is to integrate agency operations and technology investments; for instance, the electronic media, such as the Internet, is used as the means to provide a better environment for the public to comment on proposed rules and regulations. Based on the review of the received public comments, government agencies revise the proposed rules.

With the proliferation of the Internet, it becomes a growing problem for government agencies to handle the comments submitted by the public. Large amounts of electronic data, i.e., the public comments, are easily generated, and they need to be reviewed and analyzed along with the drafted rules. As such, part of e-rulemaking involves a non-trivial task of sorting through a massive volume of electronically submitted textual comments. For example,

the Federal Register (2003) documented a recent case where the Alcohol and Tobacco Tax and Trade Bureau (TTB) received over 14,000 comments in 7 months, majority of which are emails, on a flavored malt beverages proposal. The call for public comments by the TTB included the following statement:

> "All comments posted on our Web site will show the name of the commenter but will not show street addresses, telephone numbers, or e-mail addresses (2003, p. 67388)."

However, due to the "unusually large number of comments received," the Bureau announced later that it is difficult to remove all street addresses, telephone numbers and email addresses "in a timely manner (2003, p. 67388)." Instead, concerned individuals are asked to submit a request for removal of address information as opposed to the original statement posted in the call for comments. The example shows that an effortless electronic comment submission process has turned into a huge data processing problem for government agencies. Fortunately, the advance in Information and Communication Technology (ICT) can help alleviate some of the barriers in e-rulemaking. This paper will discuss a prototype of a comment analysis system, which classifies public comments according to related provisions in the drafted regulations. The automated relatedness analysis system can potentially save rule makers significant amount of time in reviewing public comments in regard to different provisions in the drafted regulations.

B A C K G R O U N D

In the field of legal informatics, most research efforts focus on enhancing the search and browse aspect of legal corpus, whose targeted users are legal practitioners. Merkl and Schweighofer (1997) suggested that "[the] exploration of document archives may be supported by organizing the various documents into taxonomies or hierarchies that have been used by lawyers for centuries (p. 465)." Indeed, a hierarchical organization of relevant public comments and drafted provisions can extend the benefit to industry designers, planners, policy-makers and interested individuals as well.

Berman and Hafner (1989) observed that legal rights of individuals are "severely compromised by the cost of legal services (p. 928)," and as a result suggested the potential of Artificial Intelligence (AI) to improve legal services. Rissland, Ashley and Loui (2003) also noted that "the law offers structure and constraints that may enable AI techniques to handle law's complexity and diversity (p. 6)." Researchers have studied extensively the application of AI, in particular, knowledge-based systems, to the *understanding* of the law (Bench-Capon, 1991; Brüninghaus & Ashley, 2001; Schweighofer, Rauber, & Dittenbach, 2001; Thomson, Huntley, Belton, Li, & Friel, 2000; Valente & Breuker, 1995; Wahlgren, 1992). In comparison, the application of ICT or AI to help the *making* of the law is rather new. Nonetheless, ICT *can* help streamline the development of regulatory policy in several new directions (Coglianese, 2003). One suggestion is to integrate rules with other laws, such as using ICT to "link all the traces of a rule's history, both back to the underlying statute and back to past or related rules,

facilitating improved understanding of legal requirements (Coglianese, 2004, p. 88)." Previous work has shown that such an application of ICT is indeed possible. A framework for comparisons among regulations from multiple sources has been developed, with successful examples of related provisions automatically linked (Lau, Law, & Wiederhold, 2003a, 2003b). This paper will exploit the use of the developed comparison framework to implement a prototype to aid the e-rulemaking process.

AUTOMATED ANALYSIS OF COMMENTS FROM E-RULEMAKING

In order to help screening and filtering of public comments, a prototype regulatory analysis system has been developed to automate the comparison between the drafted rules and their associated comments (Lau, Law, & Wiederhold, 2005). The prototyped e-rulemaking scenario incorporates public comments submitted to the US Access Board, who released a drafted chapter for the Americans with Disabilities Act Accessibility Guidelines (Access Board, 1999), titled "Guidelines for Accessible Public Rights-of-way (Access Board, 2002)." The drafted chapter was less than 15 pages long. However, over a period of four months, the Board received over 1,400 public comments which totaled around 10 Megabytes in size, with some comments longer than the draft itself. To facilitate understanding of the comments with reference to the draft, we have developed an automated system to perform a relatedness analysis.

Each piece of comment is compared with individual provisions from the draft. Characteristic features, such as conceptual phrases (Dörre, Gerstl, & Seiffert, 1999), are automatically extracted for the computation of the degree of similarity between provisions and comments (Bishop, 1995). The organizational structure of regulation is also exploited to review hidden relevance between neighboring provisions and references, similar to citation analysis and link analysis in the Web (Brin & Page, 1998; Garfield, 1995; Kleinberg, 1998; Page, Brin, Motwani, & Winograd, 1998). The results of this analysis are related pairs of provision from the draft and individual comment. Figure 1 shows the developed framework where users are given an overview of the draft along with related comments.
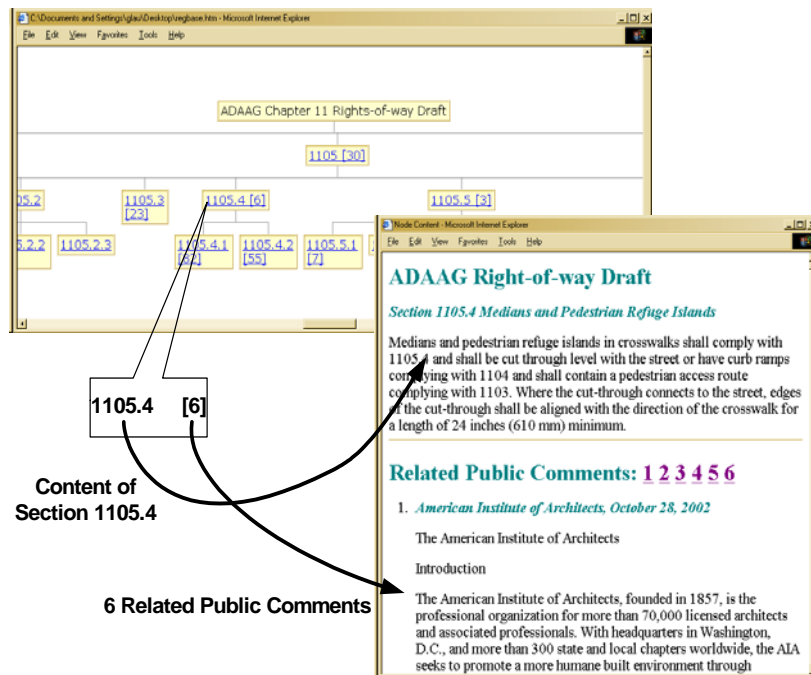

Figure 1: Comparisons of Drafted Rules and Public Comments in E-Rulemaking

As shown in Figure 1, the drafted regulation appears in its natural tree structure with each node representing sections in the draft. Next to the section number on the node, for example, Section 1105.4, is a bracketed number that shows the number of related public comments identified. Users can follow the link to view the content of the selected section in addition to its retrieved relevant public comments. This prototype demonstrates the use of a regulatory comparison system on an e-rulemaking scenario to help review drafted rules based on a large pool of public comments.

To conceptualize the use of a comparison framework on e-rulemaking, we present some interesting results here. Figure 2 shows a typical pair of drafted section and excerpts of an identified related public comment. Section 1105.4.1 in the draft established the requirements for pedestrian refuge islands in situations where there lacks adequate signal timing for full crossing of traffic lanes. Using the prototyped system, we found that one of the reviewers complained about the same situation, where in the reviewer's own words, "walk lights that are so short in duration" should be investigated. This example illustrates that our system correctly retrieves related pairs of drafted section and public comment. We also observed from this example that a full content comparison between provisions and comments is necessary, since title phrases, such as "length" in this case, are not always illustrative of the content. Automation is clearly desirable as it would otherwise require much human effort to conduct a full content comparison to sort through piles of comments.

ADAAG Chapter 11: Rights-of-way Draft

Section 1105.4.1: Length

Where **signal timing is inadequate for full crossing of all traffic lanes** or where the crossing is not signalized, cut-through medians and pedestrian refuge islands shall be 72 inches (1830 mm) minimum in length in the direction of pedestrian travel.

Public Comment, October 29, 2002

I am a member of The American Council of the Blind. I am writing to express my desire for the use of audible pedestrian traffic signals to become common practice. Traffic is becoming more and more complex, and many traffic signals are set up for the benefit of drivers rather than of pedestrians. This often means **walk lights that are so short in duration** that by the time a person who is blind realizes they have the light, the light has changed or is about to change, and they must wait …

Figure 2: Related Drafted Rule and Public Comment

The pair of highly related provision and comment shown in Figure 3 suggests that a comparison between drafted provisions and comments is indeed the right approach. This commenter started by citing Section 1109.2 in the draft, followed by a list of suggestions and questions on Section 1109.2. Our system gathered the relatedness between Section 1109.2 and this comment through different features, such as the shared phrases. This piece of comment is a representative example of a lot of comments that are written similarly: comments that are concerned about a single provision in the draft.

Thus, a comparison between drafted provisions and comments is important to help rule makers focus on the most related comments per provision.

ADAAG Chapter 11: Rights-of-way Draft

1109.2 Parallel Parking Spaces

An access aisle at least 60 inches (1525 mm) wide shall be provided at street level the full length of the parking space. The access aisle shall connect to a pedestrian access route serving the space. The access aisle shall not encroach on the vehicular travel lane. EXCEPTION: An access aisle is not required where the width of the sidewalk between the extension of the normal curb and boundary of the public right-of-way is less than 14 feet (4270 mm). When an access aisle is not provided, the parking space shall be located at the end of the block face.

Public Comment, August 26, 2002

1109.2 Parallel Parking Spaces. An access aisle at least 60 inches (1525 mm) …

1. **This section** needs to be clarified, i.e., where is the access isle located? that is, "will it be on the driver side or passenger side?"

2. The following is more of a question/concern about this requirement: …

I would really appreciate, if you could **forward this comments to the right individual** and hopefully get a response back …

Figure 3: Comment Intended for a Single Provision Only

F U T U R E   T R E N D S

Based on the observations of results shown in the previous section, there seem to be room for improvement for an e-rulemaking portal. The example

in Figure 3 shows that the public might find it helpful to submit comments on a per provision basis, in addition to a per draft basis. With enabling technology, it should be possible to develop an online submission system that allows for both types of comment submission. It saves participants time to paraphrase or cite their concerned provision. It also saves rule makers time to locate related comments either through human effort or an automated system. Comments submitted on a per draft basis can still be analyzed and compared with the entire draft to identify any additional relevant provisions. In the example of Figure 3, the commenter also suggested that it is important to forward the comments to the right person. An extension of this relatedness analysis framework could be developed to automatically inform any assigned personnel in charge of reviewing the provision within government agencies. Regulations are frequently updated by agencies to reflect environmental changes and new policies. However, the desynchronized updating of regulations seems to be problematic, especially when different regulations reference one another. We observe that there is a need for consistency check among multiple sources of regulations citing each other as references. For instance, in the domain of accessibility, Balmer (2003) pointed out that the "ADAAG references the A17.1 elevator code for conformance. Since 2000 there has been no section of the A17 that references lifts for the disabled. Therefore ADAAG references a non-existent standard … if ADAAG is to reference the A18 then the A18 should contain the requirements for this application (p. 10)." In previous work, we have developed a semi-automated

reference extraction parser, which can potentially be extended to locate cross citations and check for consistency of rulemaking (Kerrigan, 2003; Lau, 2004).

A public comment submission portal, an automated comment routing system and a reference tracking tool are just a few examples of the potential technology impact on the making of the law. E-rulemaking provides a rich research platform for a multi-disciplinary study involving social scientists, policy-makers, computer scientists, and interested and affected citizens. Improvements and innovations of an integrated e-rulemaking system are much needed to help various parties to locate, retrieve, review and validate regulations.

C O N C L U S I O N

This paper has demonstrated a potential use of ICT on policy making, in particular, the communication between government agencies and the public via comments on proposed rules. A short discussion is given on the observed impact of e-rulemaking on the efficiency of government agencies. This translates into a significant increase of workload for rule makers, as the drafted rules need to be analyzed, compared and revised based on the generated public comments. One of the main barriers to e-rulemaking, namely the vast amount of public comments received through the Internet, is concretized using a recent drafted regulation as an example. We proposed to perform a relatedness analysis on the drafted regulation and its associated public comments to streamline the process.

A prototype relatedness analysis system is demonstrated to compare a drafted regulation and its associated public comments. Documented and reported here are selected examples of results, which also inspired some potential future research directions. By screening through the public comments and sorting them according to their relatedness to provisions in the draft, it helps rule makers to review and revise the draft based on the comments. It helps interested and affected individuals to focus their concerns and suggestions to drafted rules. The use of ICT on the making of regulations can undoubtedly facilitate the editorial job of policy-makers and enhance public understanding of rulemaking.

R E F E R E N C E S

US Alcohol and Tobacco Tax and Trade Bureau (TTB). (2003). Flavored Malt Beverages and Related Proposals; Posting of Comments Received on the TTB Internet Web Site. *Federal Register, 68(231)*, 67388-67389.

US Architectural and Transportation Barriers Compliance Board (Access Board). (1999). *Americans with Disabilities Act Accessibility Guidelines (ADAAG) for Buildings and Facilities*.

US Architectural and Transportation Barriers Compliance Board (Access Board). (2002). *Draft Guidelines for Accessible Public Rights-of-Way*.

Balmer, D.C. (2003). Trends and Issues in Platform Lift. *Proceedings of Space Requirements for Wheeled Mobility Workshop*, Buffalo, NY.

Bench-Capon, T.J.M. (1991). *Knowledge Based Systems and Legal Applications*. San Diego: Academic Press Professional, Inc.

Berman, D.H., & Hafner, C.D. (1989). The Potential of Artificial Intelligence to Help Solve the Crisis in Our Legal System. *Communications of the ACM, 32(8)*, 928-938.

Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*. New York: Oxford University Press, Clarendon Press.

Bollacker, K.D., Lawrence, S., & Giles, C.L. (1998). CiteSeer: An Autonomous Web Agent for Automatic Retrieval and Identification of Interesting Publications. *Proceedings of the 2nd International Conference on Autonomous Agents*, Minneapolis, MN, 116-123.

Brin, S., & Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Proceedings of the 7th International World Wide Web Conference*, Brisbane, Australia, 107-117.

Brüninghaus, S., & Ashley, K.D. (2001). Improving the Representation of Legal Case Texts with Information Extraction Methods. *Proceedings of the 8th International Conference on Artificial Intelligence and Law (ICAIL 2001)*, St. Louis, Missouri, 42-51.

Calado, P., Ribeiro-Neto, B., Ziviani, N., Moura, E., & Silva, I. (2003). Local versus Global Link Information in the Web. *ACM Transactions on Information Systems (TOIS), 21(1)*, 42-63.

Coglianese, C. (2003). *E-Rulemaking: Information Technology and Regulatory Policy*, Technical Report, Regulatory Policy Program, Kennedy School of Government, Harvard University, Cambridge, MA.

Coglianese, C. (2004). Information Technology and Regulatory Policy. *Social Science Computer Review, 22(1)*, 85-91.

Dörre, J., Gerstl, P., & Seiffert, R. (1999). Text Mining: Finding Nuggets in Mountains of Textual Data. *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, CA, 398-401.

Garfield, E. (1995). New International Professional Society Signals the Maturing of Scientometrics and Informetrics. *The Scientist, 9(16)*, 11.

Gentner, D., & Markman, A.B. (1997). Structure Mapping in Analogy and Similarity. *American Psychologist, 52(1)*, 45-56.

Gurrin, C., & Smeaton, A.F. (1999). A Connectivity Analysis Approach to Increasing Precision in Retrieval from Hyperlinked Documents. *Proceedings of Text REtrieval Conference (TREC)*, Gaithersburg, MD.

Kepler, J. (1992, original work published 1609). *New Astronomy* (W.H. Donahue, Trans.). Cambridge, England: Cambridge University Press.

Kerrigan, S. (2003). *A Software Infrastructure for Regulatory Information Management and Compliance Assistance*, Ph.D. Thesis, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA.

Kleinberg, J. (1998). Authoritative Sources in a Hyperlinked Environment. *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms*, San Francisco, CA, 668-677.

Lau, G., Law, K., & Wiederhold, G. (2005). Legal Information Retrieval and Application to E-Rulemaking. *Proceedings of the 10th International Conference on Artificial Intelligence and Law (ICAIL 2005)*, Bologna, Italy, 146-154.

Lau, G. (2004). *A Comparative Analysis Framework for Semi-Structured Documents, with Applications to Government Regulations*, Ph.D. Thesis, Civil and Environmental Engineering, Stanford University, Stanford, CA.

Lau, G., Law, K., & Wiederhold, G. (2003a). A Framework for Regulation Comparison with Application to Accessibility Codes. *Proceedings of the National Conference on Digital Government Research*, Boston, MA, 251-254.

Lau, G., Law, K., & Wiederhold, G. (2003b). Similarity Analysis on Government Regulations. *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, DC, 111-117.

Merkl, D., & Schweighofer, E. (1997). En Route to Data Mining in Legal Text Corpora: Clustering, Neural Computation, and International Treaties. *Proceedings of the 8th International Workshop on Database and Expert Systems Applications*, Toulouse, France, 465-470.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). *The PageRank Citation Ranking: Bringing Order to the Web*, Technical Report, Stanford University, Stanford, CA.

Rissland, E.L., Ashley, K.D., & Loui, R.P. (2003). AI and Law: A Fruitful Synergy. *Artificial Intelligence, 150(1-2)*, 1-15.

Schweighofer, E., Rauber, A., & Dittenbach, M. (2001). Automatic Text Representation, Classification and Labeling in European Law. *Proceedings of the 8th International Conference on Artificial Intelligence and Law (ICAIL 2001)*, St. Louis, Missouri, 78-87.

Silva, I., Ribeiro-Neto, B., Calado, P., Moura, E., & Ziviani, N. (2000). Link-Based and Content-Based Evidential Information in a Belief Network Model. *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Athens, Greece, 96-103.

Thomson, R., Huntley, J., Belton, V., Li, F., & Friel, J. (2000). The Legal Data Refinery. *International Journal of Law and Information Technology, 8(1)*, 87-97.

Valente, A., & Breuker, J. (1995). ON-LINE: An Architecture for Modelling Legal Information. *Proceedings of the 5th International Conference on Artificial Intelligence and Law*, College Park, MD, 307-315.

Wahlgren, P. (1992). *Automation of Legal Reasoning*. Deventer, The Netherlands: Kluwer Law and Taxation Publishers.

TERMS AND DEFINITIONS

E-rulemaking: During the process of rulemaking, government agencies are required to inform and to invite the public to review a proposed rule. E-rulemaking redefines this process in which the electronic media, such as the Internet, is used to provide a better environment for the public to locate, retrieve and comment on proposed rules.

Similarity/Relatedness: Psychologists Gentner and Markman (1997) suggested that "similarity is like analogy (p. 45)," based on the concept of analogy defined by Johannes Kepler (1609/1992). In the comparative analysis of regulatory documents, we define similarity/relatedness to describe elements of texts that are alike in substance and/or connected by reason of a discoverable relation.

Relatedness Analysis: A comparative analysis among regulations and supplementary documents that identifies similar or related materials by computing a similarity score. Although the term relatedness appears more appropriate, the phrase "similarity score" has been used in the field of Information Retrieval (IR) traditionally.

Characteristic Feature: Features are the evidences that identify relatedness. Examples include conceptual phrases such as "access aisle," and domain-specific terminologies such as chemical names and symbols. Often, features are constructed by hand based on some understanding of the particular problem being tackled (Bishop, 1995).

Feature Extraction: Feature extraction is an important step in repository development when the data is voluminous. In particular, in the field of

Information Retrieval (IR), software tools exist to fulfill the task of feature extraction to recognize and classify significant vocabulary items (Bishop, 1995).

Academic Citation Analysis: It is the analysis of the citation pattern among scholar publications and literatures. For instance, CiteSeer is a scientific literature digital library that provides academic publications indexed with their citations (Bollacker, Lawrence, & Giles, 1998).

Link Analysis: The popularity of the Internet has led to extensive examinations of different types of hyperlink topology and fitting models (Calado, Ribeiro-Neto, Ziviani, Moura, & Silva, 2003; Gurrin & Smeaton, 1999; Silva, Ribeiro-Neto, Calado, Moura, & Ziviani, 2000). Google's PageRank algorithm (Brin & Page, 1998; Page, Brin, Motwani, & Winograd, 1998) and the HITS (Hypertext Induced Topic Search) algorithm (Kleinberg, 1998) both exploit the hyperlink structures of the web, similar to the heavy cross-references in the legal domain.