

A Relatedness Analysis Approach for Regulation Comparison and E-Rulemaking Applications

Gloria Lau, Haoyi Wang
Stanford University
Dept. of Civil & Environmental Eng.
Stanford, CA 94305-4020
{glau, haoyiw}@stanford.edu

Kincho Law
Stanford University
Dept. of Civil & Environmental Eng.
Stanford, CA 94305-4020
law@stanford.edu

Gio Wiederhold
Stanford University
Computer Science Dept.
Stanford, CA 94305-9040
gio@db.stanford.edu

ABSTRACT

The process of e-rulemaking with participation from the public involves a non-trivial task of sorting through and organizing a massive volume of electronically submitted comments. This research proposes to make use of available Information and Communication Technology (ICT) to help describe the relationship of public comments to policy drafts and deliberations. Based on previous work on regulatory management and comparisons, a relatedness analysis tool has been prototyped and applied to compare drafted regulations with the associated public comments. An example using a drafted regulation on rights-of-way access and the comments received by the Access Board is employed to illustrate the prototyped analysis tool. The drafted regulation and public comments are compared using not only a traditional term match but also a combination of feature matches, and not only content comparison but also structural analysis. This comparison framework helps review of comments with respect to provisions in the draft. Examples of results are shown to illustrate the use and limitations of ICT to support policy making.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *retrieval models*, I.2.1 [Artificial Intelligence]: Applications and Expert Systems – *law*.

General Terms

Algorithms, Legal Aspects.

Keywords

Relatedness Analysis, E-Rulemaking, Regulatory Comparison, Structural Analysis.

1. INTRODUCTION

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 proposed rules. Interested and affected citizens then submit comments accordingly. E-rulemaking redefines this process of rule drafting and commenting to involve the public more effectively in the making of regulations. Electronic media, such as the Internet, provide a better environment for the public to comment on proposed rules and regulations. For instance, email has become a popular communication channel for comment submission. Based on the review of public comments received in part from the electronic agora, government agencies revise the proposed rules.

The process of e-rulemaking generates a large number of public comments that need to be reviewed and analyzed along with the drafted rules. With the increased connectivity provided by the Internet, government agencies are required to handle a growing amount of data from the public. For example, the Federal Register documented a recent case where the Alcohol and Tobacco Tax and Trade Bureau received over 14,000 comments in 7 months, the majority of which were emails, on a flavored malt beverages proposal. The call for public comments 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 [13].” However, due to the “unusually large number of comments received,” the Bureau later announced that it was difficult to remove all street addresses, telephone numbers and email addresses “in a timely manner.” Instead, concerned individuals were asked to submit a request for removal of address information as opposed to the original statement posted in the call for comments. As such, an “effortless” electronic comment submission process turned into a huge data processing problem for this government agency.

As noted by Coglianese [9], Information and Communication Technology (ICT) can potentially help streamline the development of regulatory policy in several new directions. 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 [10].” Previous work has shown that such an application of ICT is indeed possible. A framework for comparisons between regulatory documents from multiple sources has been developed, with successful examples of related provisions automatically linked [23]. Based on the developed framework, this paper demonstrates another potential ICT application to support rulemaking.

This paper first reviews some literature work in Information Retrieval and different approaches to regulatory management. Section 3 discusses the technology behind a demonstrative relatedness analysis prototype that compares government regulations from different sources. Some preliminary results are given. We then show a potential application of this system to e-rulemaking to compare drafted regulations with associated public comments. We demonstrate the automated sorting of public comments with respect to drafted rules with which interested users can review related rules and comments. Rule makers also can use this tool to locate relevant public comments among thousands received. Several examples of results obtained using this tool will be shown to illustrate potential improvements as well as limitations of the use of ICT in this rulemaking scenario. Finally, observations drawn from this prototype application of comparisons between drafted rules and public comments are presented.

2. RELATED WORK

Guidance in the interpretation of government regulations has existed as long as regulatory documents. For instance, CalDAG [15] is one of many reference books written for compliance guidance with the accessibility code in California. The introduction of information technology (IT) to aid regulation exploration follows naturally. For instance, the Business Gateway¹ project, a presidential e-government initiative, aims to reduce the burden of business by making it easy to find, understand, and comply with relevant laws and regulations [28].

There are many research efforts in applying IR techniques to a legal corpus. Data mining techniques, in particular, text mining algorithms, are sought to perform automated classifications on legal documents [36]. Schweighofer et al. attempted a content-based clustering and labeling of European law, taking into account the importance of different terms [32]. Besides clustering of regulations, work has been done on improving the search experience in a legal corpus. Information extraction techniques are used to aid legal case retrieval based on a “concept” search, where “concepts” are defined to be the headnotes, heading section, case name, court name, judge, etc [26]. A similar approach is used in the SALOMON project that identified and extracted relevant information from case laws, such as keywords and summaries [25]. Finally, a natural language search capability is provided by online legal research services such as Westlaw².

Text document comparison, in particular, similarity analysis between a user query and documents in a generic corpus, is widely studied. User queries are mostly treated as a pseudo-document containing very few keywords from user input. As a result, determining the similarity between documents and user query (which can be modeled as a short document) can be modeled as document comparisons. Different techniques are developed to compute the match between user queries and documents, such as the Boolean model and the Vector model [29,

31]. Most of these techniques are bag-of-word analyses on the index terms [2]. There are a variety of algorithms to compute index term weights, and a general review can be found in [30]. We will follow a simple approach, which is to use the count of term appearance as the term weight.

In the relatedness analysis of regulations, we will introduce the notion of structural comparisons based on the hierarchical and referential organization of provisions. Due to the evolution of the Web, there has been a lot of research work related to academic citation analysis [14]. For instance, CiteSeer is a scientific literature digital library that provides academic publications indexed with their citations [3]. Different types of hyperlink topology and fitting models are examined extensively for different purposes [6, 17, 33]. While Google’s PageRank algorithm simulates web surfers’ behavior [4, 27], the HITS (Hypertext Induced Topic Search) algorithm exploits the hyperlink structures to locate authorities and hubs on the Internet [21]. In our work, the heavily referenced nature of regulations provides extra information about provisions similar to the link topology of the Web. Our domain is different from the Web - citation analysis assumes a pool of documents citing one another, whereas regulations resemble separate islands of information. Within an island of regulation, provisions are highly referenced; across islands, they are seldom cross-referenced.

3. RELATEDNESS ANALYSIS OF REGULATIONS

Starting from a well-prepared repository as described in [19, 20] and [24], we employ a combination of IR techniques and document structure analysis to extract related provisions based on a similarity measure, which is defined as a similarity score between 0 and 1. Since typical regulations are massive in size, we take a provision as the unit of comparison. Regulations are represented as trees in the analysis; thus the unit of comparison is pair of nodes in regulation trees, such as nodes A and U shown in Figure 1. The goal is to identify the most related provisions across different regulation trees using not only a traditional term match but instead a combination of feature matches, and not only content comparison but also structural analysis. This is obtained by first comparing regulations based on conceptual information as well as domain knowledge through a combination of feature matching. In addition, legal documents possess specific structures, such as the tree hierarchy of regulations and the referential structure in Figure 1. These structures also represent useful information in locating related provisions, and are therefore incorporated into the analysis for a more accurate comparison.

A base score is first computed between two sections by matching extracted features. This allows for a combination of generic features, such as concepts, as well as domain knowledge, such as measurements in accessibility regulations. This design provides the flexibility to add on features and different feature weighting schemes if domain experts desire to do so. The scoring scheme for each of the features essentially reflects how much resemblance can be inferred between the two sections based on that particular feature. For instance, concept matching is done similar to the index term matching in the Vector model [29], where the degree of similarity of documents is evaluated as the correlation between their index term vectors. Under the Vector model, a cosine similarity between the two concept vectors would represent the

¹ The Business Gateway project is formerly called the Business Compliance One-Stop project. The web address for this portal is <http://www.business.gov>.

² Westlaw online legal research service can be accessed at <http://www.westlaw.com>.

degree of similarity between the two provisions based on a concept match. Scoring schemes for other features are developed using the same idea. Some features are associated with ontologies to define synonyms, which cannot always be modeled as Boolean term matches (As an example, a domain expert can potentially define a measurement of “12 inches maximum” as 75% similar to a measurement of “12 inches”). Therefore, these feature vectors are mapped onto a different vector space before comparison to account for synonyms and non-Boolean matching [22].

The base score is subsequently refined by utilizing the tree structure of regulations. The parent, siblings and children of the interested sections are compared to include similarities between the interested sections that are not previously accounted for based on a direct comparison. Referring to Figure 1, the immediate neighbors of provision A, i.e., the parent, siblings and children, are collectively termed the $psc(A)$ of node A. In other words, similarities between the immediate neighbors imply similarity between the interested pair, which defines the basis of neighbor inclusion. The referential structure of regulations is handled in a similar manner, based on the assumption that similar sections often reference each other. Two sections referencing similar sections are more likely to be related and should have their similarity score raised. The process of reference distribution

essentially utilizes the heavily self-referenced structure of the regulation to further refine the similarity score. Figure 1 shows the out-references from provision A as the $ref(A)$ of node A. Taking Section A from the ADAAG [1] and Section U from the UFAS [35] as an example, $psc(A)$ is compared to $psc(U)$ as well as $ref(A)$ versus $ref(U)$ in score refinements. After successive refinements, similarities from both near-tree neighbors and references are identified, and related provisions are retrieved based on the resulting scores.

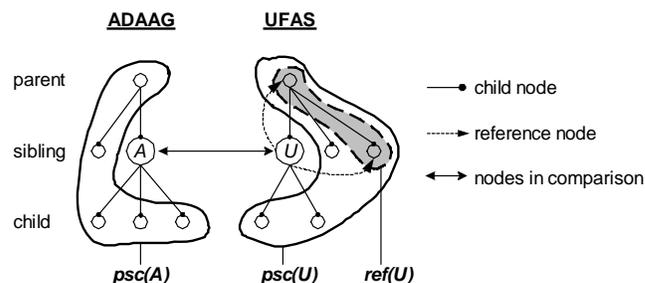


Figure 1: Immediate neighboring nodes and referenced nodes in regulation trees

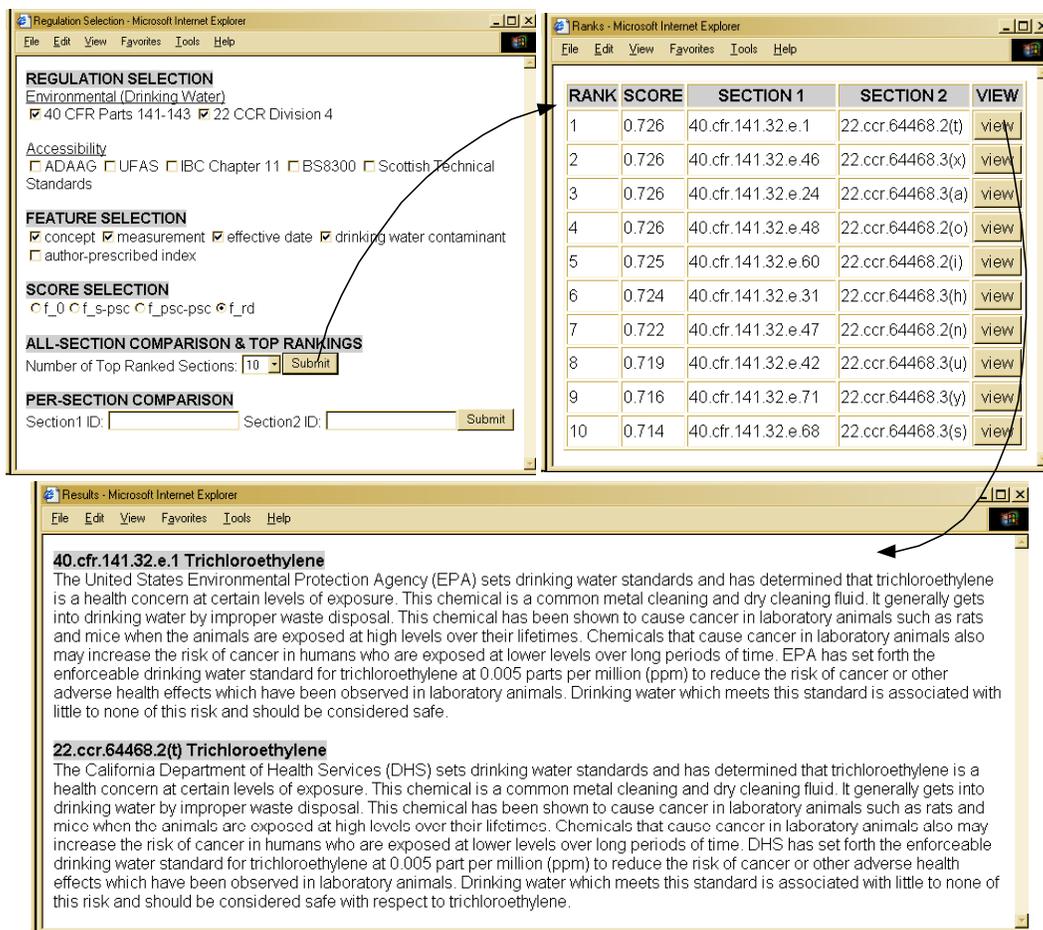


Figure 2: Ranked list of similar sections in 40CFR and 22CCR on drinking water regulations

In the domain of environmental protection, we demonstrate the prototype by comparing Federal and State drinking water regulations. Parts 141 to 143 on national drinking water standards are selected from the US Code of Federal Regulations Title 40 (40 CFR titled “Protection of the Environment”) [8], along with drinking water provisions from the California Code of Regulations Title 22 (22 CCR titled “Social Security,” Division 4 on Environmental Health) [7]. As shown in Figure 2, the relatedness analysis framework produces a ranked list of similar provisions between 40 CFR and 22 CCR on drinking water regulations as a result, based on the computation of the similarity scores.

Preliminary results obtained from the comparisons between different regulations are documented in [22]. A user survey is conducted to rank the similarity of ten randomly chosen provisions from the ADAAG [1] and ten from the UFAS [35]. The relatedness analysis system is compared with Latent Semantic Indexing (LSI) [11], as LSI claims to form concept axes instead of term axes based on Singular Value Decomposition (SVD) [16], which shares a similar goal as our feature extraction. The Root Mean Square Error (RMSE) is used to compute the ranking prediction error based on the survey results as the “correct” answer. Overall, our system outperforms the LSI with RMSE of 22.9 and 27.4 respectively. Individual combinations of features and structural matching produce prediction errors ranging from 12.0 to 29.1; majority of which are smaller than the error produced by a LSI implementation. Among the features implemented in an accessibility domain, such as concepts, measurements and author-prescribed indices, the use of measurement features results in far reduced errors such as 12.0. This reinforces our belief in domain knowledge, especially in this case, when both the ADAAG and the UFAS prescribe heavily quantified requirements that can only be captured by measurement features.

On the other hand, structural matching does not seem to affect the error in any noticeable trend. This is possibly due to the fact that the ten randomly selected pairs of provisions happen to be not very much referenced. Another explanation is that the “correct” answers do not make use of the structures either - the users are not given with much contextual and referential information in the survey for a complete understanding of the two regulations in comparison. Since this survey is only conducted using accessibility regulations, it is difficult to generalize the results to claim that the use of domain knowledge produces superior results compared to analysis performed without domain knowledge in other domains. However, the results do indicate that domain knowledge has its values in enhancing the understanding of provisions, as is apparent in the domain of accessibility based on the survey.

To justify for the proposed score refinements, we compare results obtained using the base score with results from neighbor inclusion and reference distribution. The first example shown in Figure 3 illustrates the use of neighbor inclusion, where we compare the base score with the refined score, and some improvement is observed. For instance, Section 4.1.6(3)(d) in the ADAAG [1] is concerned with doors, while Section 4.14.1 in the UFAS [35] deals with entrances. As expected, a pure concept match could not identify the relatedness between door and entrance, thus resulting

in a zero base score. However, with non-zero similarities between their neighbors, the system is able to infer some relatedness between the two sections from the neighbors in the tree. The related accessible elements, namely door and entrance, are identified indirectly through neighbor inclusions.

<p>ADAAG <i>4.1.6(3)(d) Doors</i> (i) Where it is technically infeasible to comply with clear opening width requirements of 4.13.5, a projection of 5/8 in maximum will be permitted for the latch side stop. (ii) If existing thresholds are 3/4 in high or less, and have (or are modified to have) a beveled edge on each side, they may remain.</p> <p>UFAS <i>4.14 Entrances</i> <i>4.14.1 Minimum Number</i> Entrances required to be accessible by 4.1 shall be part of an accessible route and shall comply with 4.3. Such entrances shall be connected by an accessible route to public transportation stops, to accessible parking and passenger loading zones, and to public streets or sidewalks if available (see 4.3.2(1)) ...</p>
--

Figure 3: Related provisions identified through neighbor inclusion

To illustrate the similarity between American and British standards, we compare the UFAS [35] with the BS8300 [5]. Figure 4 shows provisions from the two regulations both focusing on doors. Given the relatively high similarity score between Sections 4.13.9 of UFAS and 12.5.4.2 of BS8300, they are expected to be related, and in fact they are. Due to the differences in American and British terminologies (“door hardware” versus “door furniture”), a simple concept comparison, i.e., the base score, cannot identify the match between them. However, similarities in neighboring nodes, in particular the parent and siblings, implied a higher similarity between Section 4.13.9 of UFAS and Section 12.5.4.2 of BS8300. This example shows how structural comparison, such as neighbor inclusion, is capable of revealing hidden similarities between provisions, while a traditional term-matching scheme is inferior in this regard.

Apart from neighbor inclusion, reference distribution also contributes to revealing hidden similarities between provisions. For instance, as shown in Figure 5, both sections from the UFAS [35] and the Scottish code [34] are concerned about pedestrian ramps and stairs which are related accessible elements. However, even with neighbor inclusion, these two sections show a relatively low similarity score, which is possibly due to the fact that a pure term match does not recognize stairs and ramps as related elements. In this case, after considering reference distribution, these two provisions show a significant increase in similarity based on similar out-references. Again, this example shows how structural matching, such as reference distribution, is important in revealing hidden similarities which will be otherwise neglected in a traditional term match.

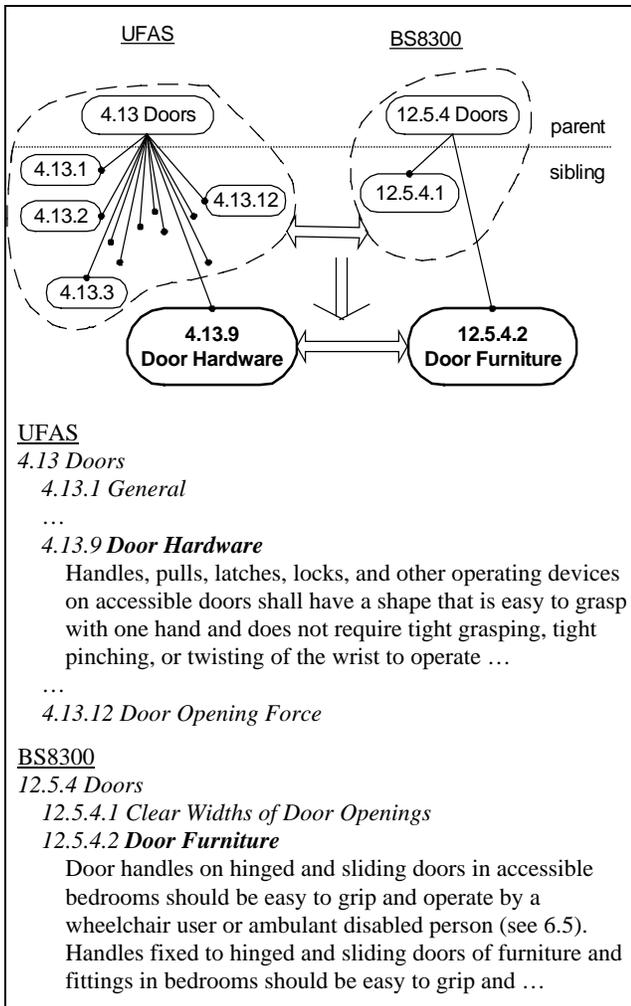


Figure 4: Example of a similarity analysis between American and British regulations

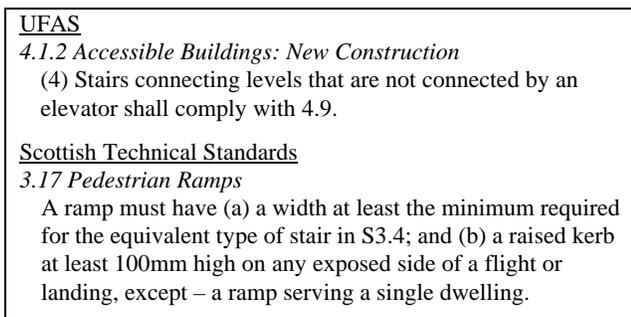


Figure 5: Related elements stair and ramp identified

4. APPLICATION ON E-RULEMAKING

Aside from the intended application on comparisons between regulatory documents, we have applied the prototype system to other domains, such as e-rulemaking, to demonstrate system scalability and extensibility. The process of e-rulemaking with participations from the public involves a non-trivial task of

sorting through a massive volume of electronically submitted textual comments. Thus, our relatedness analysis system can potentially help to screen and filter public comments. The source of data is from the US Access Board, which released a newly drafted chapter for the ADAAG [1], titled "Guidelines on Accessible Public Rights-of-way [12]." This draft is less than 15 pages long. Over a period of four months, the Board received over 1,400 public comments representing around 10 megabytes of data, where some comments are longer than the original draft. To facilitate understanding of the comments with reference to the draft, a relatedness analysis is performed on the draft chapter and the comments.

The relatedness analysis framework compares each provision from the drafted chapter with each of the 1,400 public comments. To compare provisions with comments, a similarity score is computed per pairs of provisions and comments based on the computational properties, including feature matching and structural matching as defined in the previous section. Domain-specific features, such as measurements, do not add much value here since comments coming from the general public tend to be less technical. However, commenters often follow similar terminologies found in the regulation, and therefore generic features, such as concepts, are still representative of comments. As for structural matching, we are essentially performing a single-tree (only the regulation tree but not the comments) structural analysis, since comments are not hierarchically organized. Nevertheless, neighbors and references in the draft regulation should not be overlooked.

The results of a relatedness analysis are related pairs between the provision from the draft and individual comments. Figure 6 shows the developed framework where users are given an overview of the draft along with related comments. Industry designers, planners, policy makers, as well as interested and affected individuals are potential users who can benefit from the exploration of relevant provisions and comments provided by this framework. As shown in Figure 6, 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 how a regulatory comparison system can help improve the e-rulemaking process where one needs to review drafted rules based on a large pool of public comments.

Several interesting results illustrate the potential impact as well as limitations of the use of a comparison framework on rulemaking. Figure 7 shows a typical pair consisting of drafted section and its identified related public comment. Section 1105.4.1 in the draft discusses situations in which "signal timing is inadequate for full crossing of traffic lanes." Indeed, 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, which aids user understanding of the draft. Another observation from this example is that a full content comparison between provisions and comments is necessary, because title phrases, such as "length" in this case, are not always illustrative of the content.

Automation is desirable as it would otherwise require a lot of human effort to perform a full content comparison for the large number of comments.

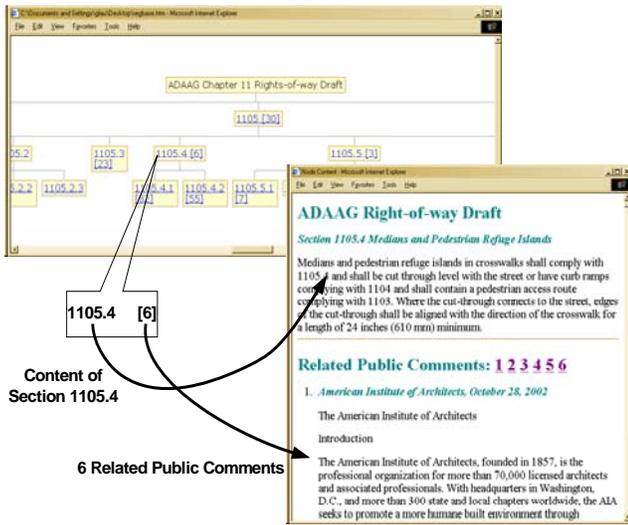


Figure 6: Comparisons of drafted rules with public comments in e-rulemaking

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
Deborah Wood, 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 for the next walk light. this situation can repeat itself again and again at such an intersection, which can make crossing such streets difficult, if not impossible. ...

Figure 7: Related drafted rule and public comment

A different type of comment screening is shown in Figure 8. It is an even more interesting result in which a particular piece of public comment is not latched with any drafted section. Indeed, this reviewer’s opinion is not shared by the draft. This reviewer commented on how a visually impaired person should practice “modern blindness skills from a good teacher” instead of relying on government installed electronic devices on streets to help. This opinion is not represented in the drafted document from the Access Board, which explains why this comment is not related to any provision according to the relatedness analysis system. As

shown in the two examples, by segmenting the pool of comments according to their relevance to individual provisions, our system can potentially save rule makers a significant amount of time reviewing public comments in regard to different provisions of the drafted regulations.

ADAAG Chapter 11 Rights-of-way Draft
 No relevant provision identified

Public Comment
Donna Ring, September 6, 2002
 If you become blind, no amount of electronics on your body or in the environment will make you safe and give back to you your freedom of movement. You have to **learn modern blindness skills from a good teacher**. You have to practice your new skills. Poor teaching cannot be solved by adding beeping lights to every big Street corner!
 I am blind myself. I travel to work in downtown Baltimore and back home every workday by myself. I go to meetings and musical events around town. I use the city bus and I walk, sometimes I take a cab or a friend drives me. Some of the blind people who work where I do are so poor at travel they can only use that lousy “mobility service” or pay a cab. Noisy street corners won’t help them.
 If you want blind people to be “safe” then pray we get better teachers of cane travel.
 I am utterly opposed to mandating beeping lights in every city. That is way too much money to spend on an unproven idea that is not even needed.

Figure 8: A piece of public comment not related to the draft

The provision and its related comment shown in Figure 9 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 about 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 users focus on the comments that are most closely related to each provision.

Based on the observation made from the example shown in Figure 9, there seems to be room for improvement for an e-rulemaking portal. The public might find it helpful to submit comments on a per provision basis, in addition to a per draft basis. With the available technology, it should be possible to develop an online submission system that allows for both types of comment submission. It saves participants the time it would take to paraphrase or cite the provision concerned. It also saves rule makers the time necessary to locate related comments either through human effort or by using an automated system. Comments submitted on a per draft basis can still be analyzed and compared with the entire draft to identify any relevant provisions. On a side note, this commenter also suggested that it is important to forward the comment to the right person. An extension of this relatedness analysis framework could be developed to inform

automatically any assigned personnel in charge of reviewing the provision within government agencies.

Apart from correctly identifying comments that are related to different provisions, limitations of our system have also been observed. Section 1109.2 is related to another comment as shown in Figure 10. The relatedness is revealed through the shared features between Section 1109.2 and the comment, which includes a direct quotation and revision of Section 1109.2. The identified relatedness is correct; however, suggested modifications and revisions of provisions cannot be detected automatically. In essence, our current system is able to uncover the relatedness but not the revised version of provisions embedded in the comments. To locate precisely the revisions suggested in the comments, one can potentially perform linguistic analysis to compute differences between the drafted version and the suggested version. This is assuming that the suggested revision does not differ significantly from the draft such that patterns can still be matched.

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

Norman Baculinao, P.E., PTOE, August 26, 2002

1109.2 Parallel Parking Spaces. An access aisle at least 60 inches (1525 mm) wide shall be provided at street level ...

1. This section needs to be clarified, i.e., where is the access aisle 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 9: Comment intended for a single provision only

Finally, Figure 11 shows a piece of public comment that is not identified as relevant to any provision in the draft. This reviewer commented on the general direction and intent of the draft, which explains why our system failed to sort this comment into any provision. Furthermore, this particular result suggests that a comparison between provisions and comments might not be enough. One could use the same analysis framework to compare comments with one another. For instance, this reviewer supported the positions of the American Council of the Blind (ACB) and the Washington Council of the Blind (WCB). While our system failed to associate this comment with any provision, comments submitted by ACB and WCB might give a clue to where this comment should belong. Essentially, clustering of

comments alone could be as handy as the illustrated clustering of comments and provisions.

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

Bruce E. Taylor, P.E., October 25, 2002

Re: Request for Comments on the Draft Guidelines for Accessible Public Rights-of-Way.

The Oklahoma Department of Transportation has reviewed the proposed draft guidelines for accessible public rights of way ...

Further, Section 1109.2, Parallel Parking Spaces, states; ...

The Department would propose that the requirements of Section 1104.12 requiring one compliant parking space per block face, be removed, and **Section 1109.2 be revised to read;**

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 or on adjacent connecting streets ...

Figure 10: Suggested revision of provision in comment

ADAAG Chapter 11 Rights-of-way Draft

No relevant provision identified

Public Comment

Douglas L. Hildie, September 13, 2002

I am responding to a request from a fellow member of the blind community in this nation. She, and I, are members of the American Council of the Blind (ACB), its state affiliate the Washington Council of the Blind (WCB), and local chapters in our communities. **I support the positions of ACB, WCB,** and many people who are blind that, failure of national, regional, and local government to provide for the require and implement rational policies and practices resulting in the installation of tactile warnings and audible pedestrian signals at intersections would be unjustified and unjustifiable ...

Figure 11: Comment on the general direction of draft

5. CONCLUSIONS AND FUTURE TASKS

E-rulemaking defines the process with which electronic media are used to aid traditional rulemaking. In particular, government agencies are required to invite public comment for newly drafted rules. Electronic media provide an easy-to-access environment for the public to submit comments. On the other hand, an increasingly unmanageable amount of electronic data, in the form of public comments, can be easily created. There is a need for an analysis tool to help rule makers and interested and affected individuals review drafted rules along with the received public comments.

Relatedness analysis combines domain knowledge with corpus-specific document structural information, such as provision hierarchy and inter-section referencing. It is shown to provide a reliable measure of similarity between pairs of provisions, based on their shared features, neighbors or references. Potential application of our system on the e-rulemaking process is demonstrated to help identify related drafted provisions and public comments. Limitations were observed, in which comments that dealt with the general intent of the drafted rules were difficult to match. It is conceivable that more pairs of "related" provisions and comments could be retrieved by relaxing the matching algorithm, such as lowering the threshold similarity score. However, this can lead to an overwhelming number of "related" matches, which might not aid in the understanding of the draft and associated public comments as intended. In addition, domain-specific features as well as structural information are not prominent among comments, making them more difficult to analyze than regulations. There is tremendous research potential to further explore the potential use of ICT for streamlining the e-rulemaking process and to provide benefits and values to both the rulemaking agencies and the public.

Potential future research directions include automated forwarding of comments to corresponding personnel in agencies, as well as automated clustering of comments. Linguistic analysis could help identify suggested provision revisions embedded in comments. An online comment submission portal, allowing for commenting per provision in addition to the existing per draft basis, could be valuable. Other applications of ICT on e-rulemaking, such as a compliance assistance system for government regulations [18], may also open routes for future research.

6. ACKNOWLEDGMENTS

This research project is sponsored by the National Science Foundation, Contract Numbers EIA-9983368 and EIA-0085998. The authors would like to acknowledge an equipment grant from Intel Corporation.

7. REFERENCES

- [1] *Americans with Disabilities Act (ADA) Accessibility Guidelines for Buildings and Facilities*, US Architectural and Transportation Barriers Compliance Board (Access Board), Washington, DC, 1999.
- [2] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*, ACM Press, New York, NY, 1999.
- [3] K.D. Bollacker, S. Lawrence and C.L. Giles. "CiteSeer: An Autonomous Web Agent for Automatic Retrieval and Identification of Interesting Publications," In *Proceedings of the 2nd International Conference on Autonomous Agents*, Minneapolis, MN, pp. 116-123, 1998.
- [4] S. Brin and L. Page. "The Anatomy of a Large-Scale Hypertextual Web Search Engine," In *Proceedings of the 7th International World Wide Web Conference*, Brisbane, Australia, pp. 107-117, 1998.
- [5] *British Standard 8300*, British Standards Institution (BSI), London, UK, 2001.
- [6] P. Calado, B. Ribeiro-Neto, N. Ziviani, E. Moura and I. Silva. "Local versus Global Link Information in the Web," *ACM Transactions on Information Systems (TOIS)*, 21 (1), pp. 42 - 63, January, 2003.
- [7] *California Code of Regulations (CCR)*, Title 22, California Office of Administrative Law, Sacramento, CA, 2003.
- [8] *Code of Federal Regulations (CFR)*, Title 40, Parts 141 - 143, US Environmental Protection Agency, Washington, DC, 2002.
- [9] C. Coglianese. *E-Rulemaking: Information Technology and Regulatory Policy*, Technical Report, Regulatory Policy Program, Kennedy School of Government, Harvard University, Cambridge, MA, 2003.
- [10] C. Coglianese. "Information Technology and Regulatory Policy," *Social Science Computer Review*, 22 (1), pp. 85-91, 2004.
- [11] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer and R. Harshman. "Indexing by Latent Semantic Analysis," *Journal of the American Society of Information Science*, 41 (6), pp. 391-407, 1990.
- [12] *Draft Guidelines for Accessible Public Rights-of-Way*, US Architectural and Transportation Barriers Compliance Board (Access Board), Washington, DC, 2002.
- [13] "Flavored Malt Beverages and Related Proposals; Posting of Comments Received on the TTB Internet Web Site," *Federal Register*, 68 (231), pp. 67388-67389, 2003.
- [14] E. Garfield. "New International Professional Society Signals the Maturing of Scientometrics and Informetrics," *The Scientist*, 9 (16), 1995.
- [15] M.P. Gibbens. *CalDAG 2000: California Disabled Accessibility Guidebook*, Builder's Book, Canoga Park, CA, 2000.
- [16] G.H. Golub and C.F. Van Loan. *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1983.
- [17] C. Gurrin and A.F. Smeaton. "A Connectivity Analysis Approach to Increasing Precision in Retrieval from Hyperlinked Documents," In *Proceedings of Text REtrieval Conference (TREC)*, Gaithersburg, MD, 1999.
- [18] S. Kerrigan. *A Software Infrastructure for Regulatory Information Management and Compliance Assistance*, Ph.D. Thesis, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA, 2003.
- [19] S. Kerrigan, C. Heenan, H. Wang, K.H. Law and G. Wiederhold. "Regulatory Information Management and Compliance Assistance," In *Proceedings of the National Conference on Digital Government Research*, Boston, MA, May 18-21, 2003.

- [20] S. Kerrigan and K. Law. "Logic-Based Regulation Compliance-Assistance," In *Proceedings of the 9th International Conference on Artificial Intelligence and Law (ICAIL 2003)*, Edinburgh, Scotland, pp. 126-135, Jun 24-28, 2003.
- [21] J. Kleinberg. "Authoritative Sources in a Hyperlinked Environment," In *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms*, San Francisco, CA, pp. 668-677, 1998.
- [22] G. Lau. *A Comparative Analysis Framework for Semi-Structured Documents, with Applications to Government Regulations*, Ph.D. Thesis, Civil and Environmental Engineering, Stanford University, Stanford, CA, 2004.
- [23] G. Lau, S. Kerrigan and K. Law. "An Information Infrastructure for Government Regulations," In *Proceedings of the 13th Workshop on Information Technology and Systems (WITS'03)*, Seattle, WA, pp. 37-42, December 13-14, 2003.
- [24] G. Lau, K. Law and G. Wiederhold. "Similarity Analysis on Government Regulations," In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, DC, pp. 111-117, August 24 - 27, 2003.
- [25] M.-F. Moens, C. Uyttendaele and J. Dumortier. "Abstracting of Legal Cases: The SALOMON Experience," In *Proceedings of the 6th International Conference on Artificial Intelligence and Law*, Melbourne, Australia, pp. 114-122, 1997.
- [26] J. Osborn and L. Sterling. "JUSTICE: A Judicial Search Tool Using Intelligent Concept Extraction," In *Proceedings of the 7th International Conference on Artificial Intelligence and Law (ICAIL 1999)*, Oslo, Norway, pp. 173-181, 1999.
- [27] L. Page, S. Brin, R. Motwani and T. Winograd. *The PageRank Citation Ranking: Bringing Order to the Web*, Technical Report, Stanford University, Stanford, CA, 1998.
- [28] *Proceedings of Business Compliance One Stop Workshop*, Small Business Administration, Queenstown, MD, July 24-26, 2002.
- [29] G. Salton. *The Smart Retrieval System - Experiments in Automatic Document Processing*, Prentice Hall, Englewood Cliffs, NJ, 1971.
- [30] G. Salton and C. Buckley. "Term-Weighting Approaches in Automatic Retrieval," *Information Processing and Management*, 24 (5), pp. 513-523, 1988.
- [31] G. Salton and M. McGill. *Introduction to Modern Information Retrieval*, McGraw-Hill, New York, NY, 1983.
- [32] E. Schweighofer, A. Rauber and M. Dittenbach. "Automatic Text Representation, Classification and Labeling in European Law," In *Proceedings of the 8th International Conference on Artificial Intelligence and Law (ICAIL 2001)*, St. Louis, Missouri, pp. 78-87, 2001.
- [33] I. Silva, B. Ribeiro-Neto, P. Calado, E. Moura and N. Ziviani. "Link-Based and Content-Based Evidential Information in a Belief Network Model," In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Athens, Greece, pp. 96-103, July 24-28, 2000.
- [34] *Technical Standards*, Scottish Executive, Edinburgh, Scotland, UK, 2001.
- [35] *Uniform Federal Accessibility Standards (UFAS)*, US Architectural and Transportation Barriers Compliance Board (Access Board), Washington, DC, 1997.
- [36] J. Zeleznikow and D. Hunter. *Building Intelligent Legal Information Systems*, Kluwer Law and Taxation Publishers, Deventer, The Netherlands, 1994.