An Information Infrastructure for Comparing Accessibility Regulations and Related Information from Multiple Sources

Gloria T. Lau, Stanford University, Stanford, CA 94305, USA (glau@stanford.edu)
Kincho Law, Stanford University, Stanford, CA 94305, USA (law@stanford.edu)

Summary

This paper describes a research project that addresses the difficulties in dealing with regulatory documents such as national and regional codes. These documents tend to be voluminous, heavily cross-referenced, possibly ambiguous and even conflicting at times. There are often multiple documents that need to be consulted and satisfied; however it is a difficult task to locate all of the relevant provisions. In addition, sections dealing with the same or similar conceptual ideas sometimes lay down conflicting requirements.

We propose a framework for regulation representation, analysis and comparison with emphasis on the extraction of similarities between provisions. We focus on accessibility regulations, whose intent is to provide the same or equivalent access to a building and its facilities for disabled persons. An XML regulatory repository is developed to extract structural as well as non-structural features from government regulations to help user understanding and computational analysis. A similarity analysis is performed between different sources of regulations. In order to achieve a better comparison between provisions, we employ a combination of feature matching and structural analysis. Results are shown on comparisons between American and European codes, as well as on the domain of electronic-rulemaking.

1 Introduction

The engineering industry is facing more and more complicated regulations and codes of practice. The complexity and diversity of regulatory documents make understanding and retrieval of regulations a non-trivial task. In particular, the existence of multiple jurisdictions, such as the Federal and state governments, leads to differences in formatting, terminology and context among regulations. There are many reference guides, that are published independent of governing bodies, attempting to help the public to better understand and comply with the regulations. As a result, the regulations, amending provisions and interpretive manuals together create a massive volume of semi-structured documents that amend, complement and potentially conflict with one another.

Curious citizens as well as industry practitioners are entitled to easy access, retrieval and comparisons of different regulations, but in reality, we lack the infrastructure as well as tools to support such kind of explorations. Productivity can be greatly increased if tools are provided to aid understanding of regulations. For instance, building designers, although more knowledgeable than the general public, have yet to search through the continuously changing provisions and locate the relevant sections related to the project, then sort through potential ambiguities in the provisions. Inspectors have to go through a similar evaluation process before a permit can be approved. Therefore, there is a need for a consolidated repository for regulatory documents such that tools can be developed to better understand and analyze regulations across different sources.

In this paper, we focus on regulations and documents related to accessibility. The intent of the accessibility regulations is to provide the same or equivalent access to a building and its facilities for disabled persons (for example, persons restricted to a wheelchair, persons with hearing and sight disabilities) and persons without qualifying disabilities. To motivate the problem, we give a classic example of such complexity and conflict found across different regulations as shown in

Figure 1 (Gibbens 2000). Both Federal and California regulations provide design requirements of a curb ramp; however, the Federal regulation (*ADAAG* 1999) focuses on wheelchair traversal, which is in conflict with the California regulation (this provision is from the 1998 version) (*CBC* 1998) focusing on the visually impaired when using a cane. The conflict is captured by the clash between the term “flush” and the measurement “1/2 inch lip beveled at 45 degrees”. Clearly, a framework for regulation analysis and comparison is much desired to alert users of related information dispersed across different sources of regulations.

<u>ADA Accessibility Guidelines</u>
4.7.2 Slope
Slopes of curb ramps shall comply with 4.8.2. The slope shall be measured as shown in Figure 11. Transitions from ramps to walks, gutters, or streets shall be flush and free of abrupt changes. Maximum slopes of adjoining gutters, road surface immediately adjacent to the curb ramp, or accessible route shall not exceed 1:20.

<u>California Building Code</u>
1127B.5.5 Beveled lip
The lower end of each curb ramp shall have a ½ inch (13mm) lip beveled at 45 degrees as a detectable way-finding edge for persons with visual impairments.

Figure 1: Federal and State regulations in direct conflict

In this paper, we describe a research prototype system that combines text mining and knowledge management techniques to help better manage, understand and analyze regulatory documents. The example domain is accessibility regulations. This paper is organized as follows: we first present the development of a legal corpus with multiple sources of regulatory documents consolidated into a unified XML format. Extraction of important features, e.g., concepts, measurements, references and so on, is also described in Section 2. Section 3 discusses the ongoing work on applying information retrieval and structural matching techniques to perform a relatedness analysis between provisions; preliminary results are shown to illustrate the identification of hidden relatedness of the provisions. Potential application of relatedness analysis for aiding the electronic-rulemaking (e-rulemaking) process is shown in Section 4. A brief summary and discussion on future works are given in Section 5.

2 Data Source and Representation

In order to develop a prototypic system, this work focuses on accessibility regulations, whose intent is to provide the same or equivalent access to a building and its facilities for disabled persons. Our corpus currently includes two Federal documents: the Americans with Disabilities Act Accessibility Guidelines (*ADAAG* 1999) and the Uniform Federal Accessibility Standards (*UFAS* 1997). In addition,

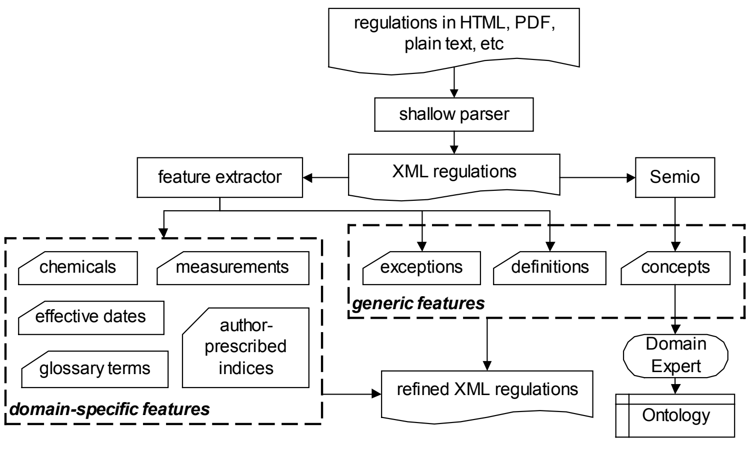

Figure 2: Repository development

Chapter 11 of the International Building Code (*IBC* 2000), titled Accessibility, is included to reflect the similarity and dissimilarity between federal and private agency mandated regulations. Related sections from the British Standard BS8300 (*British Standard* 2001) and the Scottish Technical Standards (*Technical Standards* 2001) are also included to show the differences between American and European regulations.

A brief survey on the electronic publications of regulations and supplementary documents shows that there is currently no central format for such publications. Presently, regulatory documents are available in Hypertext Markup Language (HTML), Portable Document Format (PDF) or hardcopy. To ease the development of document analysis tools, we have chosen the eXtensible Markup Language (XML) as a unified format to represent regulations in our corpus because of XML's capability to handle semi-structured data such as legal documents. To semi-automate the translational process, we have developed a shallow parser as the first phase of repository development to consolidate different formats of regulations into XML as shown in Figure 2. The structures of regulations, namely its hierarchical and referential structures, are reconstructed in XML and will be discussed in Section 2.1. Feature extraction follows in Section 2.2, where non-structural characteristics of regulations are extracted and incorporated into the XML framework. The extraction of both structural and non-structural features from regulations aims to help user understanding as well as computational analysis introduced in Section 3.

2.1 Structure Reconstruction

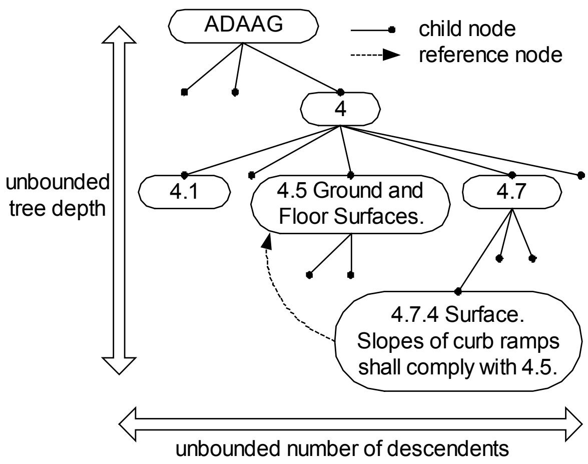

Figure 3: A regulation tree

Deep hierarchies and heavy referencing are found among provisions in regulations in general. For instance, Figure 3 shows the natural tree shape of the ADA Accessibility Guidelines (*ADAAG* 1999), where its deeply hierarchical structure is an important computational property which will be utilized in similarity analysis in Section 3. The shallow parser developed to consolidate the regulations in XML format also preserves the hierarchical structure of regulations by properly structuring provisions as XML elements. For instance, Section 4.7.4 is a provision in Section 4.7, and thus is structured to be a child node of the XML element of Section 4.7 by our shallow parser. Figure 4 shows an excerpt of the resulting XML representation of the natural hierarchy of ADAAG as shown in Figure 3.

Aside from demonstrating the hierarchy of regulations, Figure 3 also shows another important ingredient in regulations – references. The heavy referencing of regulations adds another layer of complexity in a legal corpus. In order to help user understanding as well as analysis, the shallow parser also extracts referential structures from regulations, such as the explicit reference from Section 4.7.4 to Section 4.5 in Figure 3, using tabular parsing (Kerrigan 2003). Parse trees are 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. An example of an XML reference tag is shown in Figure 4 as well, where Section 4.7.4 cites Section 4.5 once. When appropriately rendered and linked, references provide users with additional but crucial information to better understand the regulations.

```
<regulation id="adaag" name="ADA Accessibility Guidelines" type="Federal"> ...
  <regElement id="adaag.4" name="Accessible Elements and Spaces..."> ...
    <regElement id="adaag.4.7" name="Curb Ramps"> ...
      <regElement id="adaag.4.7.4" name="Surface">
        <regText> Surfaces of curb ramps shall comply with 4.5. </regText>
        <reference id="adaag.4.5" num="1" />
      </regElement> ...
    </regElement> ...
  </regElement> ...
</regulation>
```

Figure 4: XML structure of regulations

2.2 Feature Extraction

The example shown in Figure 1, where two provisions are in direct conflict, clearly demonstrates the need for a comparison system that brings together related sections in regulations. It further amplifies the importance of conceptual information, such as key phrases in the corpus (e.g., "free of abrupt changes"), as well as domain-specific information, such as measurements (e.g., ½ inch lip), for deep comparisons between provisions. However, traditional textual comparison techniques that employ simple term matching, such as the Vector model (Salton 1971), lack conceptual understanding of documents. They also suffer from the inflexibility to incorporate domain-specific information. Therefore, our comparison system, which is discussed in Section 3, combines conceptual information with domain knowledge. To enable this deeper comparison, the repository is refined with the extraction of features.

The process of feature extraction identifies the important features from the corpus that signal similarity or relatedness. As shown in Figure 2, there are two types of features: generic features that are applicable on all areas of regulations, and domain-specific features. An example of generic features is definitions, where important terms used in a regulation are defined in an early chapter of that regulation. As for domain-specific features, engineering handbooks also define in the glossary important terms used in the field, which is similar to regulation-defined definitions. For instance, the Kidder-Parker Architects' and Builders' Handbook provides an 80-page glossary that defines "technical terms, ancient and modern, used by architects, builders, and draughtsmen" (Kidder and Parker 1931). Both generic and domain-specific features represent crucial information that could potentially help user understanding and retrieval of regulations.

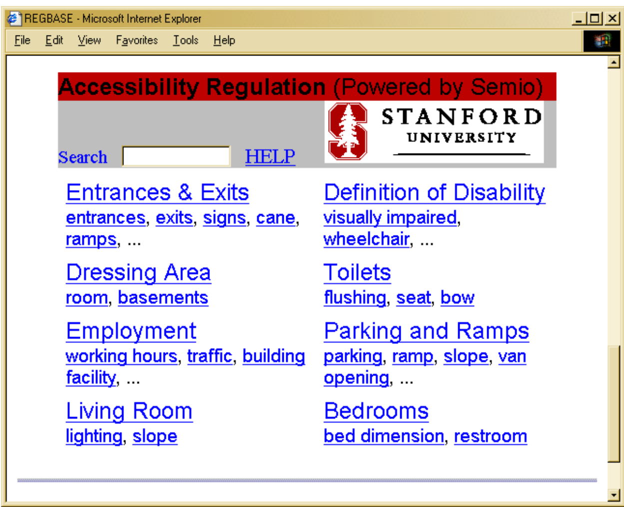

Figure 5: Concept ontology

We use a combination of text mining tool and handcrafted rules to perform a variety of feature extractions. For instance, concepts, or important noun phrases in the corpus, are identified with the help of the software tool Semio Tagger (*Semio Tagger* 2002); an example of extracted concept is the phrase "curb ramp" in Figure 1.

Figure 5 shows a potential usage of extracted features, where a concept ontology is developed by a knowledge engineer based on the list of concepts extracted, and provisions are classified automatically according to the ontology to help user retrieval of information. For other features such as measurements and definitions, handcrafted rules are implemented to automatically match them in provisions where they appear, and a frequency count is kept as well (Lau, Law and Wiederhold 2003). The corpus of documents is refined with the extracted features tagged as additional XML elements.

<u>ADA Accessibility Guidelines</u>
3.5 DEFINITIONS.
...
ACCESSIBLE.
Describes a site, building, facility, or portion thereof that complies with these guidelines.
…

<u>Refined Section 3.5 in XML format</u>
<regElement name="adaag.3.5" title="definitions">
<concept name="access aisl" num="2" />
…
<definition>
<term> **accessible** </term>
<definedAs> Describes a site, building, facility, or portion thereof that complies with these guidelines. </definedAs>
</definition>
...
</regElement>

Figure 6: XML structure of regulations with extracted definitions

<u>Uniform Federal Accessibility Standards</u>
4.6.3 PARKING SPACES.
Parking spaces for disabled people shall be **at least 96 in** (2440 mm) wide and shall have an adjacent **access aisle** 60 in (1525 mm) wide minimum (see Fig. 9). Parking access aisles shall be part of an accessible route to the building or facility entrance and shall comply with **4.3**. Two **accessible parking** spaces may share a common access aisle …
EXCEPTION: If accessible parking spaces for vans designed for handicapped persons are provided, each should have an adjacent access aisle at least ...

<u>Refined Section 4.6.3 in XML format</u>
<regElement name="ufas.4.6.3" title="parking spaces">
<concept name="**access aisl**" num="3" />
<concept name="floor surfac" num="1" />
<measurement unit="**inch**" size="**96**" quantifier="**min**" num="1" />
<reference name="ufas.4.5" num="1" />
<reference name="**ufas.4.3**" num="1" />
<index name="**access park**" num="2" />
...
<regText> Parking spaces for disabled people shall be at least 96 in ... </regText>
<exception> If accessible parking spaces for vans designed for ... </exception>
</regElement>

Figure 7: XML structure of regulations with extracted concepts, indices and measurements

Two examples of the resulting XML regulations are showed in Figures 6 and 7, each consisting of excerpts from a provision and its refined XML version. Figure 6 shows a provision in the ADAAG (*ADAAG* 1999) that defines the terminologies used in this regulation. Terms and their definitions are encapsulated in a <definition> tag, while other features, such as concepts and their frequencies of appearance in this provision, are captured in the corresponding feature tags. As shown in Figure 6, concepts are stemmed to their root forms as inflated forms are not useful in analysis. For example, the phrase "access aisle" is stemmed to "access aisl" according to Porter's Algorithm (Porter 1980). Figure 7 shows an excerpt from the UFAS (*UFAS* 1997) with several other features extracted, such as author-prescribed indices from the back of engineering handbooks, measurements and exceptions. Our system is given with a list of author-prescribed indices associated with Chapter 11 of the IBC (*IBC* 2000) and subsequently tags the provisions with these indices, which are essentially human-written concepts specific to an engineering domain. Measurements contain crucial information in the domain of disabled access as illustrated by the conflicting provisions in Figure 1, while exceptions amend the rules in the main body of provisions and should logically be separated from the body text as a separate XML element. The repository now contains both structural and non-structural information translated into a standardized XML format.

3 Automated Extraction of Related Provisions

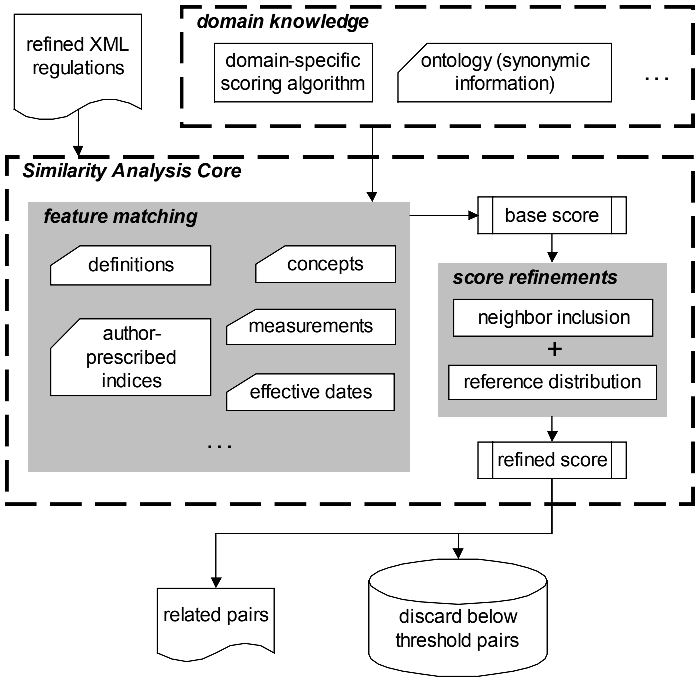

Figure 8: Similarity analysis

Starting from a well-prepared repository such as one described in Section 2, we employ a combination of information retrieval (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. A similarity score of 0 represents two provisions that are completely unrelated, whereas a score of 1 shows the highest relatedness between two provisions that are potentially identical. Since typical regulations are massive in size, we take a provision as the unit of comparison. 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 and the referential structure of regulations in Figure 3. These structures also represent useful information in locating related provisions, and are therefore incorporated into our analysis for a more accurate comparison. A schematic of our similarity analysis core is shown in Figure 8, where the details are discussed in Section 3.1 below. Section 3.2 shows some preliminary results to illustrate the use of structural information in similarity analysis.

3.1 Similarity Score Computations

A base score between two provisions is first computed by matching extracted features such as those shown in Figure 2; this allows for a combination of generic features, such as concepts, and domain knowledge, such as measurements in accessibility regulations. This design also provides the flexibility to add on features and different 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 (Salton 1971), where the degree of similarity of documents is evaluated as the correlation between their index term vectors. Using this Vector model, we take the cosine similarity between the two concept vectors as the similarity score based on a concept match.

Some features, such as field-specific glossary terms found at the back of engineering handbooks, comes with ontologies to define synonyms. Some features simply cannot be modeled as Boolean term matches due to their inherent non-Boolean property, such as measurements (e.g., a domain expert can potentially define a measurement of "12 inches maximum" as 75% similar to a measurement of "12 inches"). Some domain-specific features are provided with feature dependency information defined by knowledge experts, who do not necessarily agree with a Boolean definition. The limitation of the Vector model is observed: term axes are assumed to be mutually independent. Therefore, we modify the Vector model to accommodate dependency information, such as synonyms and non-Boolean matches, via a vector space transformation. In other words, feature vectors are mapped onto an alternate space before cosine comparisons.

In addition to incorporating domain knowledge using a combination of feature matching, our system also utilizes the hierarchical and referential structures of regulations to refine the similarity score. There are two types of score refinements: neighbor inclusion and reference distribution. In neighbor inclusion, the parent, siblings and children (the immediate neighbors) of the interested sections are compared to include similarities between the interested sections that are not previously accounted for based on a direct comparison. 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 similar sections. Essentially, reference distribution utilizes the heavily self-referenced structure of the regulation to further refine the similarity score.

The final similarity score is a linear combination of the base score, the score obtained from neighbor inclusion as well as reference distribution. We can interpret the base score as a basis of relatedness analysis formed on the shared clusters of similar features between two interested nodes, for example, Sections *A* from the ADAAG (*ADAAG* 1999) and Section *U* from the UFAS (*UFAS* 1997). Neighbor inclusion infers similarity between Sections *A* and *U* based on their shared similar clusters of neighbors in their respective regulation trees. On the other hand, reference distribution infers similarity through the shared related clusters of references from Sections *A* and *U*. In essence, the potential influence of the near neighbors are accounted for in neighbor inclusion, while the potential influence of the not-so-immediate neighbors in the tree are incorporated into the analysis through reference distribution. Thus, the final similarity score represents a combination of content comparisons between Sections *A* and *U*, and the diffusion of similarity and dissimilarity from neighboring nodes or referenced nodes to Sections *A* and *U*.

3.2 Preliminary Results of Identified Related Provisions

As a result of a similarity analysis, related provisions can be retrieved and recommended to users based on the final scores. Results obtained from the comparisons between different regulations are briefly illustrated in Figures 9 to 11 (Lau, Kerrigan and Law 2003). To demonstrate the similarity between American and British standards, we compare the UFAS (*UFAS* 1997) with the BS8300 (*British Standard* 2001). Figure 9 and Figure 10 show a subtree of 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. It is interesting to note the differences in American and British terminologies, namely "door hardware" versus "door furniture." It is also because of this terminological difference that 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.

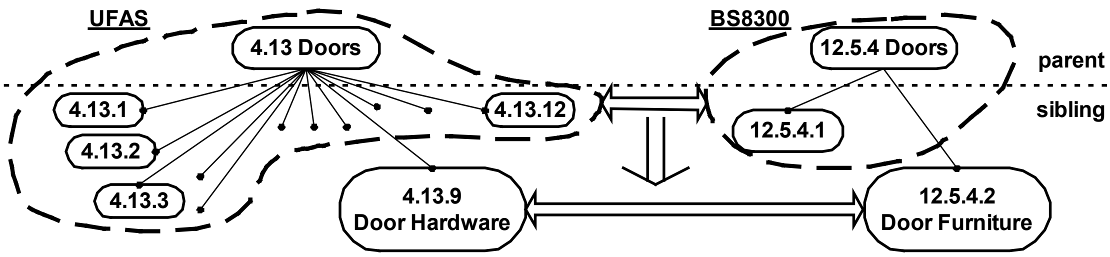

Figure 9: Similarity between neighboring nodes implies similarity between the nodes in comparison

Uniform Federal Accessibility Standards
4.13 Doors
4.13.1 General
…
4.13.9 Door Hardware
Handles, pulls, latches, locks, and other operating devices on accessible doors shall have a shape that is easy to grasp with one hand and does not require tight grasping …
…
4.13.12 Door Opening Force

British Standard 8300
12.5.4 Doors
12.5.4.1 Clear Widths of Door Openings
12.5.4.2 Door Furniture
Door handles on hinged and sliding doors in accessible bedrooms should be easy to grip and operate by a wheelchair user or ambulant disabled person …

Figure 10: Related provisions from American and British regulations

Apart from neighbor inclusion, reference distribution also contributes in revealing hidden similarities between provisions. For instance, as shown in Figure 11, both Section 4.1.2(4) from the UFAS (*UFAS* 1997) and Section 3.17 from the Scottish code (*Technical Standards* 2001) 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 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.

Uniform Federal Accessibility Standards
4.1.2 Accessible Buildings: New Construction
…
(4) Stairs connecting levels that are not connected by an elevator shall comply with 4.9.

Scottish Technical Standards
3.17 Pedestrian Ramps
A ramp must have
(a) a width at least the minimum required for the equivalent type of stair in S3.4; and
(b) a raised kerb at least 100mm high on any exposed side of a flight or landing, except – a ramp serving a single dwelling.

Figure 11: Related provisions retrieved from reference distribution

4 Application on E-Rulemaking

To demonstrate system scalability and extensibility, we have applied the prototype system to other domains as well, such as electronic-rulemaking (e-rulemaking). E-rulemaking defines the process in which the electronic media, such as the Internet, is used to provide a better environment for the public to comment on proposed rules and regulations. An example of a real scenario is as follows: the US Access Board recently released a newly drafted chapter (*Draft Guidelines for ADAAG* 2002) for the ADAAG (*ADAAG* 1999), titled "Guidelines for Accessible Public Rights-of-way." This draft is less than 15 pages long. However, over a period of four months, the Board received over 1400 public comments which total around 10 Megabytes in size. Based on the review of these public comments, the Board revises the proposed rules. As a result, the process of e-rulemaking generates a huge amount of data, i.e., the public comments, that needs to be reviewed and analyzed together with the drafted rules.

We applied our system on this domain by comparing the drafted rules with the associated public comments. Figure 12 below shows the generated output, where 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 shows how a regulatory comparison system can be very useful in an e-rulemaking situation where one needs to review drafted rules based on a large pool of public comments.

Two sample results are observed and presented here. The upper box in Figure 12 represents a typical pair of drafted section and its identified related public comment. Section 1105.4.1 discusses about inadequate signal timing for pedestrian crossing of traffic lanes. Indeed, one of the reviewers complained about the same situation that needs to be dealt with. This example illustrates that our system correctly retrieves relevant pairs of drafted section and public comment. It potentially saves rule-makers a tremendous amount of time in reviewing public comments in regard to the different provisions in the drafted regulations.

The lower box in Figure 12 shows an 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; she commented on how a visually impaired person should practice "modern blindness skills from a good teacher" instead of relying on government installment of electronic devices on the environment to help. Clearly, the opinion is not shared by the drafted document from the Access Board, which explains why this comment is not related to any provision according to our system.

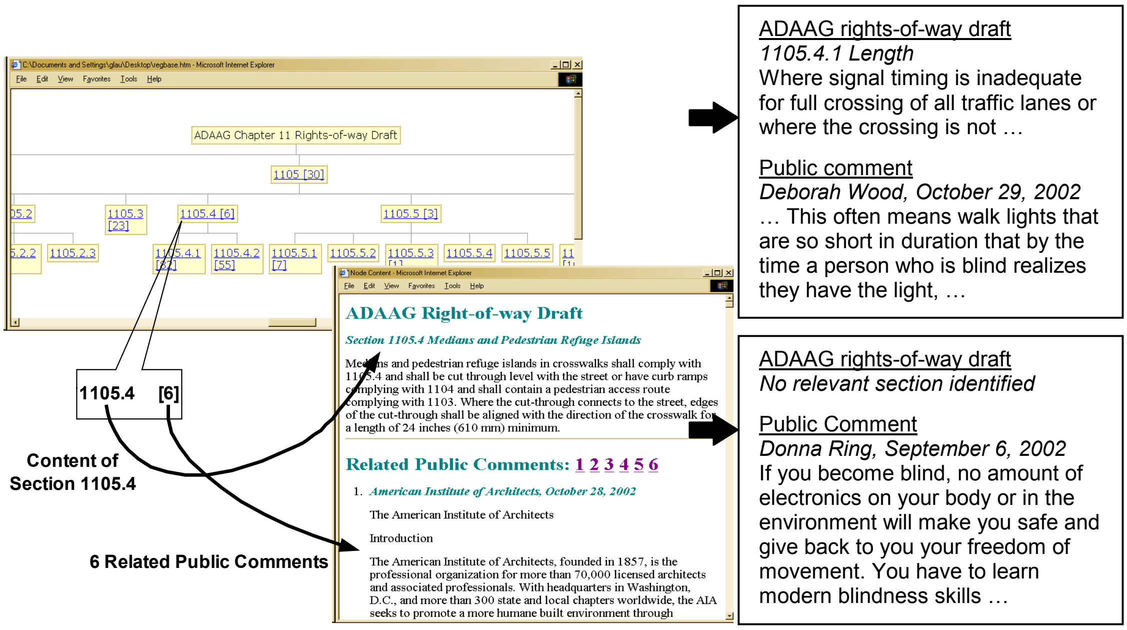

Figure 12: An e-rulemaking scenario

5 Conclusions and Future Tasks

In this paper, we present the development of a legal corpus, its associated similarity analysis, and several applications. A regulation repository is developed using XML as the standard, and our prototype includes several Federal accessibility regulations, as well as some private agency mandated standards and European codes of practice. The tree hierarchy of regulations and its referential structure are preserved by properly structuring XML elements. A combination of handcrafted rules and text mining tools is used to extract generic as well as domain-specific feature information which include concepts, measurements, definitions and so on. These features are encapsulated in XML elements whenever they appear in provisions. A similarity analysis is developed which combines IR techniques with corpus-specific document structure information. A vector space transformation is proposed to model non-Boolean domain knowledge if available. It is shown to provide a reliable measure of similarity between pairs of provisions from different regulations. Potential application of our system on the e-rulemaking process is shown to help identify related drafted provisions and public comments.

The goal of this research project is to develop an information infrastructure to aid regulation management and understanding in e-government. Due to the existence of multiple sources of regulations and the potential conflicts between them, conflict identification becomes the natural next step to a complete regulatory document analysis. We plan to study the formal representation derived from structured texts to perform an automated analysis of overlaps, completeness and conflicts.

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. We would also like to acknowledge the support by Semio Corporation in providing the software for this research.

7 References

Americans with Disabilities Act (ADA) Accessibility Guidelines for Buildings and Facilities. (1999). US Architectural and Transportation Barriers Compliance Board (Access Board), Washington, DC.

British Standard 8300. (2001). British Standards Institution (BSI), London, UK.

California Building Code (CBC). (1998). California Building Standards Commission, Sacramento, CA.

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

Gibbens, M.P. (2000). *California Disabled Accessibility Guidebook 2000*, Canoga Park, CA: Builder's Book.

International Building Code 2000. (2000). International Conference of Building Officials (ICBO), Whittier, CA.

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.

Kidder, F. and Parker, H. (1931). *Kidder-Parker Architects' and Builders' Handbook*, London, UK: John Willey & Sons.

Lau, G., Kerrigan, S. and Law, K. (2003). "An Information Infrastructure for Government Regulations," In *Proceedings of the 13th Workshop on Information Technology and Systems (WITS'03)*, December 13-14, Seattle, WA, 37-42.

Lau, G., Law, K. and Wiederhold, G. (2003). "Similarity Analysis on Government Regulations," In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 24 - 27, Washington, DC, 111-117.

Porter, M.F. (1980). "An Algorithm for Suffix Stripping," *Program: Automated Library and Information Systems 14 (3)*: 130-137.

Salton, G. (1971). *The Smart Retrieval System - Experiments in Automatic Document Processing*, Englewood Cliffs, NJ: Prentice Hall.

Semio Tagger. (2002). Semio Corporation.
http://www.semio.com.

Technical Standards. (2001). Scottish Executive, Edinburgh, Scotland, UK.

Uniform Federal Accessibility Standards (UFAS). (1997). US Architectural and Transportation Barriers Compliance Board (Access Board), Washington, DC.