A Framework for Regulation Comparison with Application to Accessibility Codes

Gloria T. Lau[1], Kincho H. Law[1] and Gio Wiederhold[2]
Department of Civil and Environmental Engineering[1], Computer Science Department[2]
Stanford University, Stanford, CA 94305
{glau,law}@stanford.edu, gio@cs.stanford.edu

Abstract

An objective of this research is to develop a framework for regulation management as well as for comparing regulations from multiple sources. An online repository for regulatory documents is created with the help of text mining tool, and users can access provisions either through the natural hierarchy of regulations or from a concept taxonomy generated by knowledge engineers. Relevant provisions are identified and brought to the user's attention through similarity analysis, which utilizes both the structure and referencing of regulations to provide a better comparison between provisions. Preliminary results show that our prototypic system is able to accurately identify similarities between Federal regulations, as well as the European codes, for disabled access.

1. Introduction

Engineering industries are facing more and more complicated regulations and codes of practice. 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.

Besides the difficulties in locating and understanding regulations, the inherent nature of multiple coding agencies also deserves attention. Regulations are typically specified by Federal as well as State governmental bodies and are amended and regulated by local counties or cities. These multiple sources of regulations sometimes compliment and modify each other, while at times one might contradict another. Designers often turn for resolution to reference handbooks that are independent of governing bodies. As a result, the regulations, amending provisions and interpretive manuals together create a massive volume of semi-structured documents with possible differences in formatting, terminology and context.

To illustrate the problem, an example of directly conflicting provisions is presented below with a section from the Americans with Disabilities Act Accessibility Guidelines (ADAAG, 1998), and a latter section from the California Building Code (CBC, 1998). Here the ADAAG focuses on wheelchair traversal while the CBC focuses on the visually impaired when using a cane. As pointed out by Gibbens (2000), "when a state or local agency requires you to construct the California required ½ inch beveled lip, they are requiring you to break the federal law," and this clearly should be brought to the user's attention.

Example 1

ADAAG
4.7.2 Slope
Slopes of curb ramps shall comply with 4.8.2. The slope shall be measured as shown in Fig. 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.

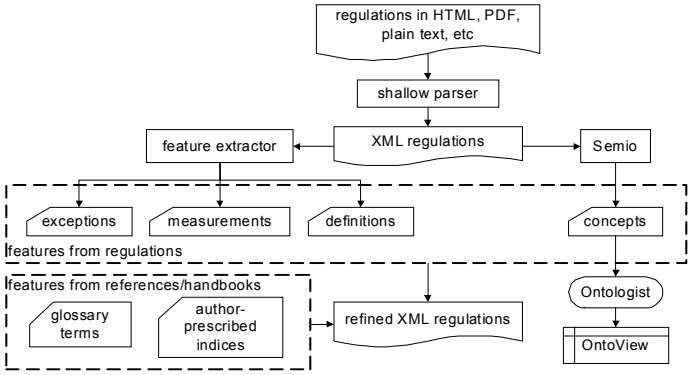
CBC
1127B.5.5 Beveled lip
The lower end of each curb ramp shall have a ½ inch (13mm) lip beveled at 45 degrees as a detectable way-finding edge for persons with visual impairments.

In this work, we propose an infrastructure for regulation comparison, and in order to develop a prototypic system, we focus 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, 1998), and the Uniform Federal Accessibility Standards (UFAS, 1986). As for comparison, selected regulations from the United Kingdom are incorporated as well, e.g. the British Standard (BS8300, 2001).

2. Data Representation

The first phase of the project is to develop an online repository of regulatory documents, and a schematic of this process is shown in Figure 1a. As for data format conversion, a shallow parser is developed to consolidate different documents into eXtensible Markup Language (XML) (W3C, 2003) for its capability to handle semi-structured data. The hierarchy of regulations is maintained by properly structuring the XML tags, e.g. Section 3.4.1 is a child node of Section 3.4, thus it is structured as a child element of Section 3.4 in the XML tree.



(a) Repository development schematic **(b) OntoView by Semio**

Figure 1. Document repository and categorization

After the documents are parsed into XML format, Semio Tagger (Semio, 2002), a commercial tool for text categorization, is used to help develop a taxonomy for regulations. A list of important noun phrases, or *concepts*, is first identified by the Tagger, with a knowledge engineer to create a taxonomy accordingly. Figure 1b above shows a taxonomy generated using Semio where users can click through the structure to view relevant provisions classified based on concepts. Repository development is complete with feature extraction, which extracts from regulations the identified features that signal related or similar sections. To illustrate the idea, an excerpt from the original provision is shown below followed by the complete set of XML mark-up of features.

Example 2

Original provision from the UFAS

4.6.3 Parking Spaces. Parking spaces for disabled people shall be at least 96 in (2440 mm) wide and shall ...
Exception: If accessible parking spaces for vans designed for handicapped persons are provided, each ...

Refined provision in XML format

```
<regElement name="ufas.4.6.3" title="parking spaces">
    <concept name="access aisle" num="3" />
    <indexTerm name="accessible circulation route" num="1" />
    <measurement unit="inch" size="96" quantifier="min" />
    <ref name="ufas.4.5" num="1" />
    <regText> Parking spaces for disabled people shall be at least 96 in (2440 mm) wide and shall ... </regText>
    <exception> If accessible parking spaces for vans designed for handicapped persons ... </exception>
</regElement>
```

In the above example, the first feature is <concept>, or key phrase extracted from the corpus by Semio Tagger. Concept helps to identify similarities and conflicts due to its ability to capture sequencing information from phrases and its simplicity compared to traditional index terms. Another source of

potentially important terms comes from author-prescribed indices (<indexTerm>) at the back of reference books or even the regulation itself; such information created manually by human can sometimes be more valuable than machine-generated phrases. Particular to accessibility provisions, measurements (<measurement>) play a very important role in which they define most of the conflicts, e.g. one provision might ask for a clear width of 10 to 12 inches, while another one might require 13 inches. Therefore, it is crucial to identify measurements and the associated quantifiers if there is any. Finally, references to other provisions and exceptions amending the body text of the provision are extracted and captured correspondingly in the <ref> and <exception> tags as shown above.

3. Similarity Comparison and Preliminary Results

We aim to develop a reliable measure of relatedness between pairs of provisions, and to recommend relevant sections of a selected provision based on the similarity measure. Here since a typical regulation can easily go over thousands of pages, we do not attempt to compare a full set of regulation against one another; rather a section from one set of regulation is compared with another section from another set, such as a comparison between Section 4.3(a) in ADAAG and Section 3.12 in UFAS. The similarity score, denoted by $f \in (0, 1)$, is derived from a base score obtained by feature matching. Feature matching, as the name suggests, matches the same features between sections; it essentially reflects how much resemblance can be inferred between the pair of sections based on that particular feature.

After obtaining the base score between pairs of sections, it is subsequently refined by taking into account the hierarchical organization of regulations as a tree structure through neighbor inclusion. Using the term *psc* to represent collectively the parent, siblings and children of a provision, i.e. its immediate neighbors, the *psc* of a section A is compared with another section B, and vice versa, to include the similarity influence of neighbors on the two interested nodes (A, B). As for the effect of the not-so-immediate neighbors of nodes A and B, reference distribution is performed; this takes into account the fact that regulations are heavily self-referenced and cross-referenced documents, and that two sections referencing similar sections are more likely to be related and should have their similarity score raised. After successive score refinements, similarities from both near-tree neighbors and not-so-immediate neighbors are accounted for, and a stable ranking of the most related sections is produced as a result.

Example 3

UFAS
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, tight pinching, or twisting of the wrist to operate. Lever-operated mechanisms, push-type mechanisms, and U-shaped handles are acceptable designs …
…
4.13.12 Door Opening Force

BS8300
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 (see 6.5). Handles fixed to hinged and sliding doors …

To assess the performance of similarity comparison, first ADAAG is compared with UFAS. Our system shows a lot of similarities between provisions from ADAAG and UFAS; indeed some of them have similarity scores of 1. This correctly reflects the fact that the two federal codes are very similar, and that some sections are even identical. To illustrate the effect of structural comparison between nodes in the regulation tree, and also the difference between American and British codes, UFAS is compared with BS8300. Example 3 shows sections from the UFAS and BS8300 both focusing on doors. Given the

relatively high similarity score (0.425) 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; Section 4.13.9 from the American code is titled "Door Hardware" while Section 12.5.4.2 from the British standard is titled "Door Furniture". As the American and British phrasing is different, a direct term match cannot identify the similarity between "door hardware" and "door furniture"; however, by comparing the neighbors of the sections, a higher similarity score is observed (0.471). As shown in Figure 2, similarities in neighboring nodes in the regulation trees imply a higher similarity between the sections compared.

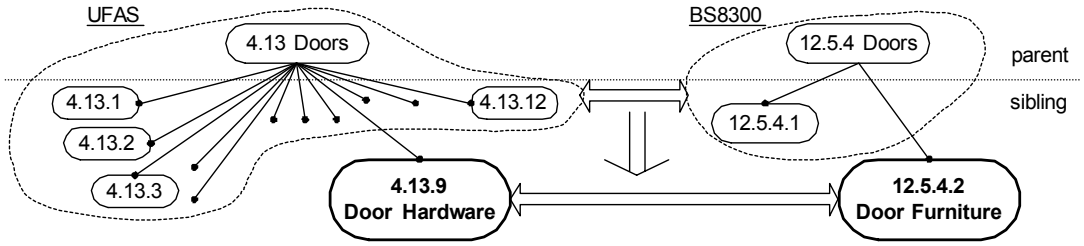


Figure 2. Score refinement based on neighboring nodes in tree

4. Conclusions and Future Tasks

The objective of this research is to develop an infrastructure for regulation comparison with current focus on accessibility codes. A repository is built by transforming regulations into XML format because of its capability to handle semi-structured data. After all regulations are in a unified format, features, or evidences, are extracted from the corpus automatically or by a knowledge engineer, in addition to features from reference materials such as engineering handbooks. Similarity analysis is then performed by first computing a base score between pairs of provisions using feature matching. Neighbor inclusion and reference distribution are performed to further refine the scores according to the structure of the regulations. A list of the most related sections is produced as a result.

Preliminary results are obtained by comparing several sets of accessibility regulations, and examples are given to show that our system does reveal relatedness between provisions through neighbor inclusion and reference distribution. To demonstrate scalability and practicality of the system, we anticipate the incorporation of other regulations in the near future once the prototype is thoroughly tested on accessibility regulations. In addition, due to the existence of multiple sources of regulations and thus potential conflicts between them, conflict identification becomes the natural next step for regulatory document analysis.

5. Acknowledgments

This research project is sponsored by the National Science Foundation, Contract Numbers EIA-9983368 and EIA-0085998. The authors would like to acknowledge a "Technology for Education 2000" equipment grant from Intel Corporation. We would also like to acknowledge the support by Semio Corporation in providing the software for this research.

6. References

(ADAAG 1998) *ADA Accessibility Guidelines for Buildings and Facilities*, The Access Board, 1998.
(BS8300 2001) *British Standard 8300*, British Standards Institution, 2001.
(CBC 1998) *California Building Code*, California Building Standards Commission, 1998.
(Gibbens 2000) Michael P. Gibbens, *California Disabled Accessibility Guidebook 2000*, Builder's Book, Canoga Park, CA, 2000.
(Semio 2002) *Semio Tagger*, Semio Corporation, 2002.
(UFAS 1986) *Uniform Federal Accessibility Standards*, The Access Board, 1986.
(W3C 2003) World Wide Web Consortium, *eXtensible Markup Language (XML)*, 2003.