Exercising Due Diligence in Legal Requirements Acquisition:
A Tool-supported, Frame-based Approach

Travis D. Breaux
North Carolina State University
tdbreaux@ncsu.edu

Abstract

Government laws and organizational policies introduce critical legal requirements that govern information systems. Unlike traditional requirements elicited from stakeholders, legal requirements have unique characteristics that software engineers must address to ensure that their systems are demonstrably compliant with relevant laws and policies. This paper presents important terminology for developing legally compliant software systems and a methodology consisting of procedures and models for acquiring, representing and analyzing phenomena in legal documents, which constitute rich sources of legal requirements. Based on a grounded theory, the method has been validated through a mixed-methods approach consisting of multiple, descriptive case studies. This paper presents a human subject experiment that tests a fundamental part of the theory to understand the efficacy of multiple users applying the method to a sample regulation text.

1. Introduction

Software developers must ensure that their software systems exhibit *legal compliance*, which includes "efforts to demonstrate a defensible position in a court of law" [5]. Although legal compliance can only be tested and proven in a court of law, developers can take legally established, pre-emptive measures that include documenting evidence of *due diligence*, which means "reasonable efforts that persons make to satisfy legal requirements", and *good faith*, which means "honesty in belief or purpose, faithfulness to one's duty or obligation" [11]. Documentation of these efforts must be integrated into the software engineering process, beginning with legal and software requirements.

This paper presents experimental results of the Frame-Based Requirements Analysis Method (FBRAM), a rigorous, systematic method for acquiring legal requirements from U.S. federal regulations with an emphasis on recording evidence of due diligence and good faith. The contributions of this paper include: (1) an overview of a critical procedure in the method that is used to identify legal requirements in the law; and (2) results from a preliminary experimental evaluation of the method using human subjects.

The remainder of the paper is organized as follows: Section 2 presents related work; Section 3 presents an overview of the method and brief discussion of theoretical foundations upon which the method builds; Section 4 presents the experimental design; Section 5 presents the experimental results; and Section 6 concludes with the discussion.

2. Related Work

Early efforts to model legislation and statutory law identified important challenges to legal requirements acquisition, including: legal models should match the language and structure of legal documents [17, 18] and normalize expressions of the logical structure [1, 3]; models should map between acquired logical propositions and corresponding paragraphs in the law [17, 18]; and models should syntactically and semantically explicate ambiguity [3]. The FBRAM, partially presented herein, addresses each of these challenges.

Research in legal software requirements acquisition is relatively recent. Breaux et al. discovered several techniques to identify rights and obligations in U.S. federal regulations [6, 9] and online privacy policies [4]. In addition, the role of legal document structure and the need for traceability between this structure and acquired requirements has been identified as a key concern [9, 12]; this observation is missing from traditional views about traceability [13]. Kiyavitskaya et al. have since used natural language processing to codify legal text with concepts of rights and obligations with partial coverage [15], demonstrating the need for a method that combines tool-support with human intervention. The FBRAM, presented herein, extends this work with new procedures and models in a new study using tool-support. This paper also reports results from a human subject experiment to evaluate one part of the FBRAM. For a survey of legal requirements beyond requirements acquisition, see Otto and Antón [16].

3. The Frame-Based Method

The Frame-Based Requirements Analysis Method (FBRAM) includes a manual, tool-supported annotation process that is used to transform a legal document into a discrete set of legal requirements. The method automatically maintains traceability between the requirements and the original legal text using a model of

IEEE
computer society

the paragraph structure of the law, a markup language and associated tool-based parser.

Engineers use concepts from an upper ontology (see Figure 1), phrase heuristics and a context-free markup language to assign an interpretation to a legal text. While the validity of this interpretation must be checked by legal experts, performing these techniques provides engineers and legal counsel a critical first step in exercising due diligence by staged coordination.

The statement-level legal concepts are defined below. Note that an *entity* is any stakeholder, system or component, including software or hardware.

- *Permission* (P) means any state that an entity is permitted to achieve, maintain or avoid, or any act that an entity is permitted to perform; permissions include stakeholder *rights*.
- *Obligation* (O) means any state that an entity is required to achieve or maintain, or any act that an entity is required to perform.
- *Refrainment* (R) means any state that an entity is required to avoid, or any act that an entity is required not to perform.
- *Exclusion* (E) means any state that an entity is not permitted or required to achieve, maintain or avoid, or any act that an entity is not permitted or required to perform.
- *Fact* (F) means any state or act that is occurring or has occurred in the past.
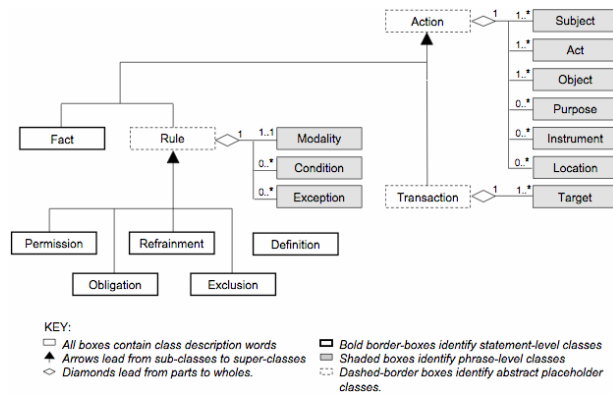- *Definition* (D) means a statement of the meaning of a specialized word, such as a term-of-art.

Figure 1: Legal requirements upper ontology

Several statement-level concepts identified in these studies correspond to the Hohfeld legal concepts for strict rights (or claims) and duties, which are similar to the above permission and obligation definitions, respectively [14]. Hohfeld identifies respective opposites that he calls no-rights and privileges. The no-right corresponds to a refrainment, sometimes called a prohibition. Hohfeld argues that "not obligated" implies a privilege, which is distinct and separate from permis-

sion. In our upper ontology, privileges correspond to exclusions and are treated equally with other acts that are not permitted, prohibited or required.

Table 1 presents concept codes that are used in the markup in Figure 2 to align sentences and phrases with concepts in the upper ontology from Figure 1.

Table 1: Upper ontology concept codes

Code	Concept	Code	Concept
P	Permission	a	Act
O	Obligation	o	Object
E	Exclusion	t	Target
s	Subject	c	Condition
m	Modality	e	Exception

Figure 2 presents an excerpt from HIPAA Privacy Rule §164.520(a)(2)(ii) with the markup in **bold** and phrases matching the phrase heuristics from Table 2 in gray highlight. The raw text from this excerpt, excluding the markup and highlighting, was used in the experiment (see Section 4). The markup is used to structure legal text into two types of nested blocks denoted by opening and closing brackets: (1) *pattern blocks*, denoted by curly "{}" brackets, indicate the start of a requirements natural language pattern or sub-pattern; and (2) *value blocks*, denoted by square "[ ]" brackets, indicate spans of text that will be mapped to slot values in a frame-based requirement [7]. A block is *typed*, if the opening bracket is followed by the concept code operator "#" and a unique concept code (see Table 1). Within any block, the English conjunctions "and" and "or" are mapped to logical connectives using the operators "&" and "|" for logical-and and logical-or, respectively (e.g., lines 4 and 16).

```
1    (ii) {#O [#s A group health plan [that
2    provides health benefits solely through
3    an insurance contract with a health
4    insurance issuer or HMO, &and that
5    [creates |or receives] [protected
6    health information in addition to
7    summary health information as defined
8    in 164.504(a) |or information on
9    whether the individual [is
10   participating in the group health plan,
11   |or [is enrolled in |or has disenrolled
12   from] a [health insurance issuer |or
13   HMO] offered by the plan]]], [#m
14   must]:{
15       (A) [#a Maintain] [#o a notice
16   under this section]; |and
17       (B) [#a Provide] [#o such notice]
18   {#c upon [request]} {#t to [any
19   person]}}}.
```

Figure 2: Example legal text annotated with markup

Official paragraph indices in U.S. federal regulations can refer to multiple requirements, with sub-paragraph indices referring to subsets of such requirements. To maintain traceability, the FBRAM includes a document model that maps acquired legal require-

ments to official indices and titles for parts, sections and paragraphs [7]. For example, the document model maps the cross-reference §164.520(a)(2)(ii) to the two requirements "to maintain a notice" (lines 1-16) and "to provide a notice" (lines 1-14, 17-19), whereas reference §164.520(a)(2)(ii)(A) to the sub-paragraph only refers to the first of these two requirements. The case studies reveal that the document model is critical to assist engineers with integrating constraints on requirements that are cross-referenced from other paragraphs.

The markup language has an existing tool-based parser, used to conduct the replication study and human subject experiment, that highlights phrase heuristics, detects syntax and semantic errors—missing brackets and unknown concept codes—and alerts the engineer who must then resolve these errors.

4. Experimental Design

The experiment investigates the following hypothesis: knowledge of the upper ontology concepts by engineers is sufficient to consistently identify legal requirements from a legal text. To test this hypothesis, we measure the level of participant agreement in a classification task, wherein participants assign upper ontology concepts to statements in a legal text using the markup language. We are interested in identifying ontological gaps (e.g. statements with no assigned concepts) and sources of measurable disagreement. Significant disagreement reinforces the need to employ phrase heuristics to automatically identify, and have engineers confirm, the appropriate classification for legal statements. This section describes the means of analysis, participant population, recruitment method, and the procedure used to conduct this experiment.

4.1. Means of Analysis

Two analysis techniques were used to analyze the markup application data: a qualitative signal analysis technique developed for this study and a quantitative inter-rater reliability statistic. The signal analysis technique treats each concept in the upper ontology as a "signal" that spans the length of the sample text. The rater agreement for a concept assigned to each word corresponds to the signal strength. For example, consider this hypothetical scenario for two raters:

Rater 1: `[#P An individual has a right to no-`
`tice...]`
Rater 2: `An individual [#P has a right to] no-`
`tice...`

Rater 1 classifies the entire statement as a permission "P", whereas Rater 2 restricts this concept to the phrase "has a right to". Figure 3 illustrates the rater agreement using the signal analysis technique for hypothetical raters 1 and 2: the upper ontology concepts appear along the y-axis, including no concept was as-

signed to a word (N), permission (P) and obligation (O). The x-axis corresponds to the word progression in the legal text from start to end; the numbers and major ticks indicate the end of the numbered statement. The minor ticks and gray box are not part of the analysis, but serve to distinguish the classifications for each word, highlighting the second word "individual" across signals N, P, and O for this example only.
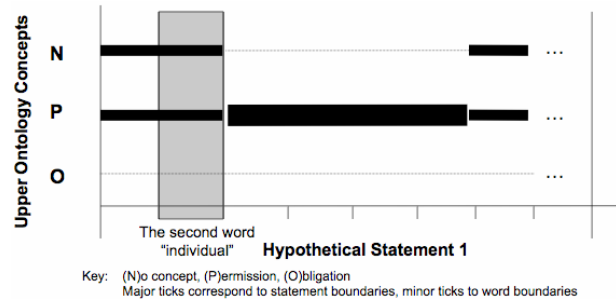

Figure 3: Example illustrating waveform analysis

For each word, the "strength" of a concept is portrayed along the x-axis as expanding vertically one increment for each rater that classifies the word with that concept. For example, the second word "individual" is highlighted in Figure 3, illustrating that one rater did not classify this word (N) and the other rater classifies it as a permission (P). Conversely, both raters agree on the next four words, "has a right to", and the signal strength of the permission signal (P) increases whereas the no concept signal (N) decreases. Because raters can assign upper ontology concepts to individual words using the markup, the signal analysis technique provides a qualitative, visual indication of potential sources of disagreement. These indications are used to formulate new hypotheses, as discussed in Section 4.2.

The participant concept classifications constitute nominal data and thus require a relevant statistic for analysis. Cohen's Kappa statistic for two raters is used to measure agreement in nominal data [10]. The Cohen's Kappa statistic yields a value [0,1] and measures actual observed agreement among two raters (P_a) excluding the expected agreement (P_e), if agreement were due strictly to chance, and is expressed by the formula: $\kappa = (P_a - P_e) / (1 - P_e)$.

In this experiment, Cohen's Kappa is used to measure agreement between a single participant's classifications and the desired outcome, consisting of classifications determined by applying the phrase heuristics, which are unambiguous for two significant U.S. federal regulations analyzed in two case studies; these include the sample text used in this experiment.

4.2. Participant Population and Recruitment

The ideal participant population consists of requirements engineers in small, medium and large en-

terprises. These individuals are few in number and typically have 10-20 years of industry experience [2, 19], which makes recruitment cost-prohibitive and impractical for many cash-strapped, government-funded research projects. However, this experiment measures the process of legal requirements *acquisition* and not *refinement*; the former requires knowledge to recognize and format requirements, whereas the latter requires considerable domain knowledge to create new requirements. As a reasonable alternative to hardened industry experts, we recruited graduate students enrolled in a 16-week requirements engineering seminar course. These students were assigned to read 32 prominent academic and industry papers on a broad spread of requirements topics, including requirements elicitation, acquisition, specification, traceability and ambiguity, prior to participating in this experiment.

The classification task was conducted as an in-class assignment wherein students' may anonymously "opt-in" to having their anonymized coursework used for research purposes. The demographic breakdown for the 12 participants in this sample appears in Table 2. We note the participants are relatively younger (interquartile range = 5.3 years) and less experienced (interquartile range = 4.4 years) than the ideal population [19] and a majority (83%) of students spoke English as their first language (English L1), a relevant factor to analyzing English legal texts.

Table 2: Participant demographic summary

Demographic	Measure
Doctoral Student	58%
Masters Student	42%
English L1	83%
Male	67%
Female	33%
Age, yrs (mean/median)	24.9/ 23.5
Industry Experience, yrs (mean/median)	3.3/ 2.3

4.3. Procedure

The experimental procedure includes four steps, listed below with the maximum time (in minutes) that was required to complete each step.

1. (1 min) The participant reads, electronically signs and receives a copy of the informed consent form, which introduces the purpose of the experiment and their rights under the Institutional Review Board (IRB), as required by the U.S. National Science Foundation under 45 CFR §§690.101-.124.
2. (7 min) The participant studies a tutorial, which introduces the terminology and requirements acquisition task that they will perform in step 3.
3. (30 min) The participant performs the classification task to acquire legal requirements from a sample legal text.

4. (2 min) The participant completes a demographic survey to measure the variables listed in Table 5.

In step #2, the two-page tutorial presents: (a) the verbatim definitions of statement-level and phrase-level concepts in the upper ontology (see Section 3); and (b) example markup for a single statement, including logical operators (see Section 3). The sample legal text in step #3 consists of HIPAA §164.520(a).

5. Experimental Results

After the participant's applied the markup to the sample text, the markup was analyzed using the signal analysis technique and Cohen's Kappa statistic, both described in Section 4.1. Figures 4, 5 and 6 present the signal analysis results. Along the y-axis appears the statement-level classifications: no concept (N), permission (P), obligation (O), refrainment (R) and exclusion (E) from the upper ontology in Section 4.2; along the x-axis appears the progression of words from the sample text's five statements.

Figure 4 illustrates, in gray highlight, the rater agreement with the phrase heuristics. The matching concepts appear across the top of the figure along the x-axis. The sample legal text contains no refrainments (R). This figure illustrates multiple sources of rater disagreement with the phrase heuristics, which is now discussed using Figures 5 and 6, including the effects of modal verbs, cross-references and section titles.
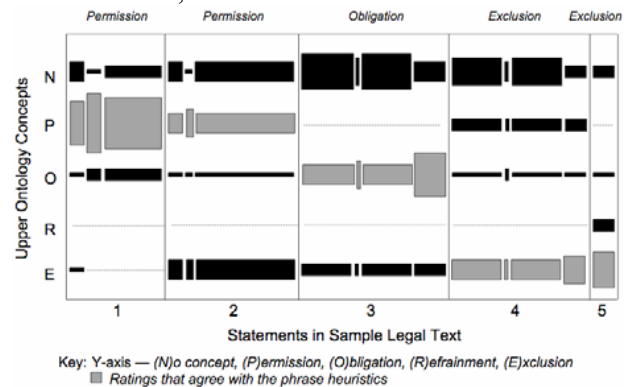

Figure 4: Rater agreement with phrase heuristics

Figure 5 highlights intra-statement anomalies in which some raters are observed changing their concept classification from "no concept" to one of permission, obligation or exclusion within a single statement, indicated by gray highlight. This effect occurs at modal verb phrases from known phrase heuristics, which appear across the top of Figure 5. A less frequent anomaly occurs where one rater changes from "no concept" to obligation when a cross-reference appears in statements #3 and #4. Both cross-references consist of the phrase "as defined in 164.504(a)", which may cause the rater to feel obliged to follow the cross-reference.
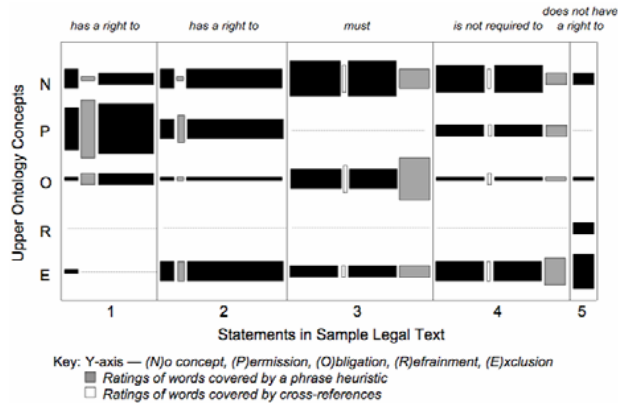
Figure 5: Intra-statement anomalies

Figure 6 highlights several transitions, in which rater disagreement shifts with changes in the section titles, which appear above the figure along the x-axis. For example, we observe a minority of raters assigned the exclusion concept to statement #2, despite the statements' modal verb phrase "has a right", which indicates a permission. The corresponding section title contains the word "exception", which may have confused some raters. This effect appears to diminish by statement #3 but returns in statement #4. The section titles may influence the classifications, as observed for statements #1 and #5, which have titles "right to notice" and "exception for inmates" that precede permissions and exclusions, respectively.
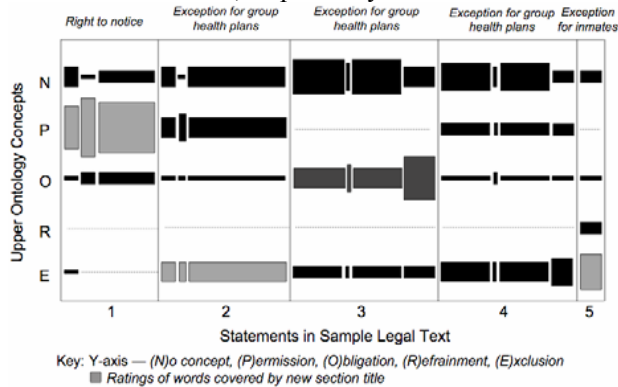

Figure 6: Rater agreement and section titles

Figure 7 presents a histogram for Cohen's Kappa for two raters: each participant's ratings compared to the ratings determined by the phrase heuristics, which constitute the desired outcome. Whereas histograms are normally presented as bar graphs, this histogram is presented as line graph to compare Cohen's Kappa for three distinct agreement measures: the complete text, the text without section titles, and only the modal verb phrases (see discussions of Figures 5 and 6). An expert rater with familiarity with the method (scoring $\kappa = 1.0$) was excluded from the figure. Mean agreement across the three measures is low at 29.6%, 27.9% and 40.6%,

respectively, with the highest single rater agreement at 63.2%, 63.9% and 73.2%, respectively, which are all within the lower bound of desired agreement, 60-80%.
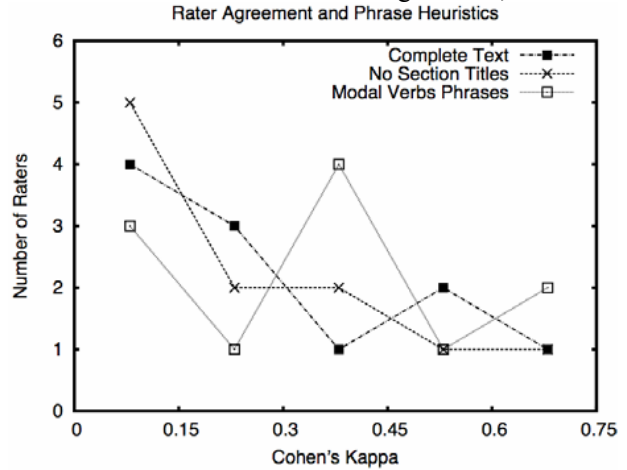

Figure 7: Cohen's Kappa for rater and heuristics

The Wilcoxon matched-pairs signed-rank non-parametric test is used to test whether the difference in mean agreements is significant, because the distribution for the inter-rater agreement is non-normal, the sample size is small ($n = 11$) and the sample pairs for each rater across the three measures are related [22]. Compared to the complete text, the text without section titles exhibits a significant decrease in agreement ($p < .005$) and the modal verb phrases exhibit a significant increase in agreement ($p < .021$). The decreased agreement may be due to raters confusing whether section titles constitute legal statements.

In summary, the experimental results indicate that the upper ontology concepts (and the Hohfeld legal concepts) are insufficient heuristics to yield significant agreement among multiple raters. Consequently, if unambiguous phrase heuristics are employed in authoring regulations, the phrase heuristics overcome this limitation in legal requirements acquisition. Moreover, the impact of section titles and cross-references on requirements acquisition was only indirectly observed and requires further study.

6. Discussion and Summary

This paper presents preliminary, experimental evaluation of the Frame-Based Requirements Analysis Method (FBRAM) for acquiring legal requirements, which is based on a grounded theory discovered through multiple case studies in two domains, information privacy and accessibility. The evaluation presented herein indicates that the FBRAM's upper ontology concepts and Hohfeld legal concepts alone are not sufficient to acquire legal requirements from regulations. Thus, the phrase heuristics are an integral part to legal requirements acquisition. By standardizing legal lan-

guage, which has been observed in two prominent U.S. federal regulations, the phrase heuristics may yield a classification of legal statements that is consistent across domains.

For the broader challenge of aligning software requirements with relevant laws, the FBRAM is only a partial solution as evidenced by eight requirement refinement patterns, which provide reasonable explanations for the kinds of gaps that exist between legal and software requirements [8]. In addition to these patterns, we envision incorporating the FBRAM legal requirements into other requirements acquisition activities to contextualize legal requirements in the context of ongoing business practices.

A future study is planned to contrast traditional requirements specification practices [20, 21] with the FBRAM to study the effects of the ontology and phrase heuristics. We hypothesize that the traditional practices, by emphasizing the "shalls" that correspond to obligations, would cause an analyst to overlook important permissions, exclusions and possibly refrainments found in legal texts. This future study also seeks to measure analysts' ability to classify legal statements using the upper ontology independent of the FBRAM. Additional research to investigate ways to improve comprehension of legal requirements by a technical audience is also envisioned.

7. Acknowledgements

8. References

[1] L.E. Allen, C.S. Saxon. "Computer aided normalizing and unpacking: Some interesting machine-processable transformations of legal rules", *Computing Power and Legal Reasoning*, pp. 495–572, St. Paul, Minnesota, 1984. West Publishing Company.

[2] J. Aranda, S. Easterbrook, G. Wilson. "Requirements in the wild: how small companies do it", *IEEE 15th Int'l Req'ts Engr. Conf.*, Delhi, India, pp. 39-48, 2007.

[3] C. Biagioli, P. Mariani, D. Tiscornia. "ESPLEX: A rule and conceptual model for representing statutes", *1st Int'l Conf. on AI and Law*, Boston, MA, pp 240–251, 1987.

[4] T.D. Breaux, A.I. Antón. "Analyzing goal semantics for rights, permissions and obligations", *IEEE 13th Int'l Req'ts Engr. Conf.*, Paris, France, pp. 177–186, 2005.

[5] T.D. Breaux, A.I. Antón, C-M. Karat, J. Karat. "Enforceabilty vs. accountability in electronic policies", *IEEE 7th Workshop on Pol. Dist. Sys.*, London, Ontario, pp. 227-230, 2007.

[6] T.D. Breaux, A.I. Antón. "Analyzing regulatory rules for privacy and security requirements", *IEEE Trans. Soft. Engr.*, 34(1), pp. 5-20, 2008.

[7] T.D. Breaux, A.I. Antón. "A systematic method for acquiring regulatory requirements", *6th Int'l Workshop on Req'ts for High Assurance Sys.*, Delhi, India, 2008.

[8] T.D. Breaux, A.I. Antón, K. Boucher, M. Dorfman. "Legal requirements, compliance and practice: an industry case study in accessibility", (In Press) *IEEE 16th Intl Req'ts Engr. Conf.*, Barcelona, Spain, 2008.

[9] T.D. Breaux, M.W. Vail, A.I. Antón. "Towards compliance: Extracting rights and obligations to align requirements with regulations", *IEEE 14th Int'l Req'ts Engr. Conf.*, Minneapolis, MN, pp. 46–55, 2006.

[10] J. Cohen. "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, 20:37–46, 1960.

[11] B.A. Garner, ed. *Blacks Law Dictionary, 8th ed.* ThompsonWest, St. Paul, MN, 2004.

[12] S. Ghanavati, D, Amyot, L. Peyton. "Towards a framework for tracking legal compliance in healthcare", *19th Int. Conf. on Advanced Information Systems Engineering*, Trondheim, Norway, LNCS v. 4495, 218-232, 2007.

[13] O. Gotel, A. Finkelstein. "Contribution structures", *2nd IEEE Int'l Symp. on Req'ts Engr.*, San Diego, California, pp. 100–107, 1995.

[14] W.N. Hohfeld. "Some fundamental legal conceptions as applied in judicial reasoning", *The Yale Law Journal,* 23(1):16–59, 1913.

[15] N. Kiyavitskaya, N. Zeni, L. Mich, T.D. Breaux, A.I. Antón, J. Mylopoulos. "Extracting rights and obligations from regulations: towards a tool-supported process", *IEEE/ACM 22nd Int'l Conf. Auto. Soft. Engr.*, pp. 429-432, 2007.

[16] P.N. Otto, A.I. Antón. "Addressing legal requirements in requirements engineering," *IEEE 15th Int'l Req'ts Engr. Conf.*, Delhi, India, pp. 5-14, 2007.

[17] M.J. Sergot, F. Sadri, R.A. Kowalski, F. Kriwaczek, P. Hammond, H.T. Cory. "The British Nationality Act as a logic program", *Communications of the ACM*, 29(5):370–386, 1986.

[18] D.M. Sherman. "A Prolog model of the Income Tax Act of Canada", *1st Int'l Conference on AI and Law*, Boston, MA, pp. 127–136, 1987.

[19] S. Sim, T. Alspaugh, B. Al-Ani. "Marginal notes on amethodical requirements engineering: what experts learned from experience", (In Press) *IEEE 16th Int'l Req'ts Engr. Conf.*, Barcelona, Spain, 2008.

[20] I. Sommerville and P. Sawyer. *Requirements Engineering: A good practice guide*. John Wiley and Sons, Ltd., West Sussex, England, 1997.

[21] R.J. Wieringa. *Requirements Engineering: Frameworks for Understanding*. John Wiley and Sons, Ltd., West Sussex, England, 1996.

[22] F. Wilcoxon, "Individual comparisons by ranking methods", *Biometrics Bulletin*, 1(6): 80-83, 1945.