Annotating Regulations Using Cerno: an Application to Italian Documents - Extended Abstract

Nicola Zeni
Nadzeya Kiyavitskaya
Dept. of Information and
Communication Technology
University of Trento, Italy
{nadzeya, nzeni}@dit.unitn.it

James R. Cordy
School of Computing
Queens University, Kingston,
Canada
cordy@cs.queensu.ca

Luisa Mich
Dept. of Computer and
Management Sciences
University of Trento, Italy
luisa.mich@unitn.it

John Mylopoulos
Dept. of Information and
Communication Technology
University of Trento, Italy
jm@science.unitn.it

ABSTRACT

The increasing complexity of software systems and growing demand for regulations compliance require effective methods and tools to support requirements analysts activities. Internationalization of information systems due to both economics and Web based architectures call for the application of regulations written in different languages. Thus far existing approaches for extracting rights and obligations have concentrated on English documents. In this paper, we describe the results of the application of Cerno, a lightweight framework for semantic annotation, to legal documents written in Italian. In addition, we investigate critical issues for semantic annotation tools in a different cultural and environmental context. Results obtained, while preliminary, allow us to quantify the effort needed to port tools based on Cerno and give some insight on directions of future development of a multilingual system to support semantic annotation of regulations not only in different domains, but also written in different languages.

Categories and Subject Descriptors

D.2.1 [**Requirements/Specifications**]: Tools; I.7.5 [**Document and Text Processing**]: Document Capture—*Document analysis*; K.5.2 [**Legal Aspects of Computing**]: Governmental Issues—*Regulation*

General Terms

Legal Aspects, Management, Experimentation

Keywords

regulation compliance, privacy requirements, tool support

1. INTRODUCTION

Increased competition and globalization of businesses coupled with new legislation in many countries is forcing organizations to continuously undertake detailed evaluations of their businesses. The evaluation must assess not only how well a business performs, but also how the productivity and profitability can be improved through the introduction of new business processes or the modification of existing ones, while leveraging the investment they have made in their existing, such as enterprise-resource-planning (ERP) systems, customer relationship management(CRM), or supply change management (SCM).

The adaptation of legacy software related to existing solutions of an organization so that it is compliant with legislated regulations of different countries is extremely expensive in terms of compliance verification and implementation.

This phenomenon and related challenges are considered by *e-Procurement* that is the business-to-business or business-to-consumer purchase through the Internet in the context of internationalization. Thus, extracting requirements from regulations is a major challenge in need of methodological aids and tools. To help meet these expectations, many people work towards providing automated solutions for ensuring regulation compliance. In order to verify if a system is compliant with a regulation, one must first identify the requirements imposed by the regulatory document.

A methodology for extracting requirements from regulations is presented in [8] using a process called Semantic Parametrization. Accordingly, the process for extracting requirements from regulations consists of three steps:

- regulatory text is annotated to identify text fragments describing actors, rights, obligations, etc.;

- a semantic model is constructed from these annotations; and

- the semantic model is transformed into a set of functional and nonfunctional requirements.

This methodology was taken as a basis of the tool that extracts rights and obligations [16] and had been preliminary evaluated on the HIPAA Privacy Rule [14]. The tool is based on the semantic annotation framework Cerno [18].

In this paper, we present the definition of a process for the extraction of rights and obligations from regulatory documents written in

the Italian language. The purpose of this study is to verify the generality of the Cerno-based annotation method, and eventually improve its effectiveness. The regulation document we considered is the Stanca Law [5], which describes accessibility requirements that must be respected by all Web sites of the Italian Public Administration in order to assure accessibility for the disabled. The document consists of a series of technical requirements and general restrictions that web sites must respect. We discuss specific problems for the annotation tool caused by the change of language, and present an experimental evaluation of the new process on the Stanca Law.

The contributions of the present paper include: (1) investigation of specific problems for the annotation tool caused by the change of language; (2) adaptation of the Cerno-based regulation analysis method to Italian documents; (3) an experimental evaluation of the new process on the Stanca Law.

The paper is structured as follows: Section 2 elaborates the language specific problems found when working with the Stanca law; following the identified challenges Section 3 presents the Cerno-based process for extracting rights and obligations adapted for Italian documents; the evaluation results for the Stanca law are presented in Section 4; related work is discussed in Section 5; finally conclusions are drawn in Section 6.

2. THE LANGUAGE DEPENDENT ASPECTS OF LEGAL DOCUMENTS

Changing language in the semantic annotation process, e.g., Japanese, Italian, Arabic, and others, means that there are at least two distinct levels to consider: (a) syntactic dimension, (b) semantic dimension. The first level is well investigated in the area of Information Retrieval technologies and basically comprises the following issues:

- Segmentation of text into words in different languages;

- Words stemming and morphology;

- Identification of phrase boundaries, i.e., punctuation conventions.

There are many tools for handling these issues, however often sofware developers underestimate the value of syntactic aspects, while a successfull text analysis tool should always cater for them.

The greatest difficulty is caused by the semantic level because it is linked to the variety of ways of world perception and cultural background. In summary, the main problems related to this level are: proper names, temporal and spatial conventions; language patterns, i.e., specific ways to express certain concepts; representation of co-reference, time reference, place reference; causal relationships, associative relationships. Not stopping on the semantic issues in detail, we would like to emphasize the importance of awareness of these aspects when designing a text analysis tool.

As regards the specificity of legal documents, first of all they are characterized by a higly structured format. In the Italian legal system there are nine levels of the document hierarchy (see Fig. 1), where the mimimun unit is so called *comma*. Comma identifies the minimum length of a meaningful phrase in the text and also is used for cross-referencing norms in citation. Document hierarchy can also help in understanding the structure of sentences. Thus, correct disambiguation of structural units is of great importance.

Moreover, standards, policies and regulation are written in a specialized language called *legalese*, which makes acquiring requirements from regulations a challenging task [10]. Legalese contains heavily qualified phrases that are laden with ambiguities, a pervasive phenomenon with natural languages in general [6].

```
Libro [numero ordinale]
 [Testo rubrica]
Parte [numero ordinale]
  [Testo rubrica]
  Titolo [numero ordinale]
  [Testo rubrica]
  Capo [numero ordinale]
  [Testo rubrica]
  Sezione [numero ordinale]
  [Testo rubrica]
  Art.  [numero cardinale arabo progressivo
all'interno di tutto l'articolato] ( [Testo rubrica] )
   1.  [testo comma]:
    a) [testo lettera]:
      1) [testo numero];
      2) [testo numero];
    b) [testo lettera].
   2.  [testo comma].
```

Figure 1: The general structure model in the Italian legal system
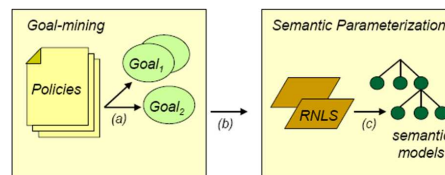

Figure 2: Manual methodology

Last but not least, one of the major challenges in analysis of policies and regulations is caused by their *prescriptive* nature [19]. A prescriptive text according to deontic logic, is caracterized by the fact that we can assign on each sentence a worth value. In a legal document each sentence can contain a prescription to do or not do something. This means that in prescriptive documents, unlike in descriptive ones, stakeholders cannot afford to overlook regulatory requirements. A higher precision and recall for annotation or text-mining is therefore required in this domain. Missing an exception or conditional information could imply the application of a right or of an obligation to the wrong set of actors or stakeholders. The cost of such error is great. Therefore, an automatic tool's accuracy should be always assessed by a comprehensive evaluation of the results.

3. CERNO-BASED PROCESS

The tool-supported process for regulation analysis that we previously developed is based on the methodology for extracting stakeholder requirements from regulations by Breaux et al. [10], see Fig. 2. In this methodology, requirements engineers mark regulatory text using phrase heuristics [9], [10] to identify rights or obligations, and associated constraints. Then, semi-formal rights, obligations and constraints are formally modeled in first-order predicate logic using a process called Semantic Parameterization that provides increased precision [8], [7]. After that, the semantic models can be analyzed for inconsistencies and corrected by an expert. In the end, a set of requirements is produced from these models.

The tool we developed for the HIPAA rule recognizes document structure in terms of section and subsection boundaries, titles and annotated paragraph indices, identifies instances of the concepts actor, policy, event, information and date and annotates document fragments describing *rights, anti-rights, obligations, anti-obligations,* and related *constraints*. To generate these annotations, the tool used a list of normative phrases for the objects of concern
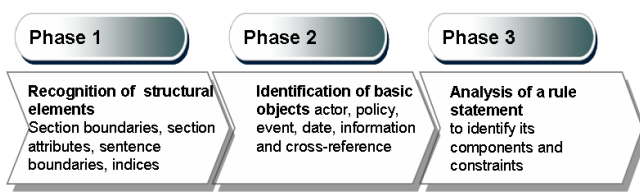
Figure 3: Cerno-based regulation analysis

Concept type	Indicators
Right	tenses of "potere"
Anti-Right	tenses of "non potere"
Obligation	tenses of "dovere", "obbligare"
Anti-Obligation	tenses of "non dovere", "non obbligare"
Constraint	"in caso di", "entro"

Table 1: Normative phrases for Italian laws

that was obtained by manual analysis of the HIPAA document [10].

Some of the indicators are complex patterns which combine both literal phrases and general concepts. The identified normative phrases assume a preliminary recognition of the following basic constructs: *cross-reference* can be of two types: internal references that refer the reader of a regulation to another paragraph within the regulation and external references, a citation of another regulation, act or law; *policy* can be the name of the law, standard, act or other regulation document which establishes rights and obligations; *actor* can be an individual or an organization involved. To recognize these objects, we extended the Parse step of Cerno's framework with the corresponding object grammars.

Our regulation analysis process consists of three main phases [16], as shown in Fig. 3:

- Recognition of structural elements of the document: section boundaries, section attributes which are number and title, sentence boundaries;

- Identification of basic objects: actor, policy, event, date, information and cross-reference;

- Deconstruction of a rule statement to identify its components and constraints.

The method is based on Cerno, a semantic annotation framework [18]. The process for generating semantic annotations in Cerno is based on a "design recovery" process borrowed from software reverse engineering [12]. This process uses a series of successive transformation steps. Cerno's framework has been applied and evaluated for semantic annotation of text documents in the following two domains: announcements for hotels and related accommodation taken from on-line newspapers, and Tourist Board web sites [17], [18].

3.1 Adapting Cerno for Italian regulations

The change of language of legal documents affects all the steps of the process. One first adaptation consist in dealing with the document structure. The grammar that we used for HIPAA application was adapted for the Italian legal document structure, as discussed in Section 2.

Another aspect relates to the parsing of special characters, as for instance letters with accent and apostrophes. Therefore, we extended words grammar in Cerno adding such specific symbols as acceptable word tokens.

In Cerno's semantic annotation process, the domain model plays an important role in guiding the analysis. It expresses primary input expectations in terms of language structures and provides a basis for identifying its concepts. It is important to note that, in order to be adaptable to different applications, Cerno factors out reusable domain-independent components and domain-dependent knowledge that can be easily modified. The annotation of English legal documents was done for the HIPAA Privacy Rule [14] according to the methodology developed by Breaux and Antón [8],

[7]. Using manual analysis of the HIPAA document, a list of normative phrases was developed that identifies many of these objects of concern [10]. The concepts used for annotation in both past and present applications are: *actor*, *policy*, *obligation*, *right*, *constraint*, *resource*, and *action*, where:

- A *right* is an action that a stakeholder is conditionally permitted to perform.

- An *obligation* is an action that a stakeholder is conditionally required to perform.

- In contrast, *anti-rights* and *anti-obligations* state that a right or obligation does not exist.

- A *constraint phrase* is the part of a rule statement that describes a single pre-condition.

To recognize instances of the Actor and Policy concepts, we can exploit the regularity of the document, meaning that as in most regulations and policies, the Stanca law uses standard terms and limits the use of synonyms to the definitions of those terms. Normally, each regulation or policy document contains an introduction section, where every used term is strictly defined and a reference synonym is assigned. However, for identification of actor instances, we adapted two solutions: (1) some instances were mined manually from the definition section "Definizioni"; (2) in order to catch the actors not mentioned in the definitions, we exploited the results provided by a Part of Speech Tagger (POS) [22], i.e., all proper nouns we marked as actors. For resource instances, we followed only the first solution reusing the terms stated in the definition section.

In order to identify action verbs, we adapted the following heuristic: annotate all verbs in present tense, passive tense and impersonal tense. The verbs in the listed forms also refer to obligations, in accordance with the instructions for writing Italian legal documents [13]. Thus, the corresponding heuristic rule was adapted for identifying obligations.

As for rights, obligations and their antis, it is more difficult to identify them in Italian language. Unlike in English language, that mainly uses modal verbs to state prescriptions as for instance " the users **must** present their request", Italian regulations normally use present active ("gli utenti presentano la domanda"), present passive ("la domanda è presentata") and impersonal tenses ("la domanda si presenta") of verbs to describe an obligation. The choice of the style highly depends on a lawmaker. Each of these styles is equally recommended by the law writing guidelines [13].

Therefore, in identification of these concepts, our strategy included (1) translation of normative phrases identified by Breaux and Antón (see the list of equivalent normative phrases in Fig. 1), (2) in addition, annotation of those sentences that contain verbs in the tenses that intrinsically express obligations as instances of obligation.

Fig. 4 shows a subset of the grammar for syntactic indicators integrated as a domain-dependent component of Cerno which cor-

```
Action:   acced[ere], adegu[are], alleg[are],
effettu[are], gest[ire], impegn[are],
individu[are]...;
Resource:   compit[o|i], ret[e|i], immagin[e|i],
indicator[e|i],indirizz[o|i], informazion[e|i]...;
Actor:   Presidente della Repubblica, Repubblica,
Amministrazion[e|i], Autorità, Cnipa, Comunità,
disabil[e|i], Ministr[o|i], Organizzazion[e|i],
valutator[e|i]...;
Obligation:   dov[ere], è fatto obbligo, farla
osservare, promuov[ere], comport[are], costituiscono
motivo di preferenza, defin[ire];
AntiObligation:   non dov[ere], non sia, non si
applica, non si possono stipulare, non esprim[ere];
Right:   po[sso|uoi|uò|ssiamo|tete|ssono|ssa];
AntiRight:   non po[sso|uoi|uò|ssiamo|tete|ssono|ssa];
```

Figure 4: A fragment of entity that we used for identify categories

```
Art.  10
(Regolamento di attuazione)
<Obligation>
1.  <Entro novanta giorni dalla data
di entrata in vigore della presente
<Policy>legge</Policy></Constraint>,
con <Policy>regolamento</Policy> emanato
ai sensi dell'articolo 17, comma 1, della
<Policy>legge</Policy> 23 agosto 1988, n.  400, sono
definiti:
a) i criteri e i princpi operativi e organizzativi
generali per l'accessibilità;
b) i <Resource>contenuti</Resource> di cui
all'articolo 6, comma 2;
c) i controlli esercitabili sugli operatori privati
che hanno reso nota l'accessibilità dei propri siti
e delle proprie <Resource>applicazioni</Resource>
informatiche;
d) i controlli esercitabili sui <Actor>soggetti</Actor>
di cui all'articolo 3, comma 1.
2.  Il <Policy>regolamento</Policy> di cui al
comma 1 è adottato previa consultazione con le
associazioni delle <Actor>persone disabili</Actor>
maggiormente rappresentative, con le associazioni di
sviluppatori competenti in materia di accessibilità
e di produttori di <Resource>hardware</Resource> e
<Resource>software</Resource> e previa acquisizione
del parere delle competenti Commissioni parlamentari,
<Constraint>che <Action>devono</Action>
pronunciarsi entro quarantacinque giorni dalla
richiesta</Constraint>, e d'intesa con la
Conferenza unificata di cui all'articolo 8 del
<Policy>decreto</Policy> legislativo 28 agosto 1997,
n.  281.  </Obligation>
```

Figure 5: A fragment of the annotated Stanca law

respond to various concepts. Symbol "|" is used to list alternative endings for the parser.

As a result, the regulation analysis process for the Stanca law includes the following steps:

- Pre-process the document by normalizing text: removing leading characters and trailing spaces;

- Identify the structure of the document, i.e., indices of the document hierarchy, and identify sentence boundaries;

- Identify and mark up basic entities: actor, policy, resource and action;

- Annotate text fragments related to the concepts of interest: right, obligation, their antis, and associated constraints.

See a fragment of the annotated Stanca document in Fig. 5.

Table 2: Quantitative evaluation summary for the Stanca law

	Actor	Action	Resource	Policy	O	AO	R	AR	C
Cerno	241	77	279	86	26	2	7	1	12
Human	170	55	58	3	24	2	9	0	32

4.   EMPIRICAL EVALUATION

After extending the Cerno-based process with the knowledge specific for Italian language, we applied it to the full text of the Stanca law. The automatic annotation of the Stanca law, containing a total of 6185 words (280 lines), by the Cerno framework takes only 61 milliseconds on a personal computer Intel Pentium 4, 3 GHz processor, RAM 2 Gb, running Suse Linux. As a result, a total of 683 basic entities and 36 rights and obligations were identified.

Table 2 presents the quantitative results of the evaluation showing the number of instances of the concepts of interest the tool identified compared to a human annotator. In this table, "R" stands for right, "O" for obligation, "AR" for anti-right, "AO" for anti-obligation, "C" for constraint.

The tool outperformed the human marker in identification of instances of the concepts actor, policy, action, and resource. This outcome leads us to conclusion that the tool was able to largely support humans in identification of relevant information.

As for complex concepts, the tool identified nearly all instances of rights and obligations, however the performance was essentially lower for constraint concept. This fact is caused by the lack of normative phrases for their reliable identification. Out plan of future work includes further investigation of how this drawback can be resolved.

There were also particular difficulties of Italian text for both human and tool that emerged. For instance, frequently the subject is omitted, as in passive forms of verbs, or hidden by using impersonal expressions, thus making it difficult to correctly classify the whole text regulatory fragment and find the holder of a right of an obligation. In comparison, the official English translation of the Stanca law in most of the cases explicitly states this information:

- Italian version: *"Nelle procedure svolte dai soggetti di cui all'articolo 3, comma 1, per l'acquisto di beni e per la fornitura di servizi informatici, i requisiti di accessibilità stabiliti con il decreto di cui all'articolo 11 costituiscono motivo di preferenza a parità di ogni altra condizione nella valutazione dell'offerta tecnica, tenuto conto della destinazione del bene o del servizio. La mancata considerazione dei requisiti di accessibilità o l'eventuale acquisizione di beni o fornitura di servizi non accessibili adeguatamente motivata."*

- English version: *"The subjects mentioned in article 3, when carrying out procedures to buy goods and to deliver services, are obliged, in the event that they are adjudicating bidders which all have submitted similar offers, to give preference to the bidder which offers the best compliance with the accessibility requirements provided for by the decree mentioned in article 11."*

As the fragment shows, in the English version the translator disambiguated implicit information.

However, the results of this annotation provide a useful input for software engineers looking for requirements contained in the regulation, rather than they start from scratch.

On overall, the results suggest that the Cerno-based process for regulation analysis is applicable to documents in written in different languages. The effort required to adapt the framework for the

new application was relatively small as regards the implementation. This application allowed us to further generalize and automate the Cerno framework in terms of modularity and parameterization.

5. RELATED WORK

The idea of using contextual patterns or keywords to identify relevant information in prescriptive documents is not new. A number of methodologies based on similar techniques have been developed. However, tools to realize and synthesize these methods under a single framework are lacking. In this section we provide a survey of existing methods designed for the analysis of regulatory documents with different levels of automation.

In [11], the authors suggested an algorithm for detection and classification of non-functional requirements (NFRs). In a pilot experiment, the indicator terms were mined from catalogs of operationalization methods for security and performance softgoal interdependency graphs. These indicators were then used to identify NFRs in fifteen requirements specifications. The results have shown a satisfactory recall and precision for the security and performance keywords.

Antón proposed the Goal-Based Requirements Acquisition Methodology (GBRAM) to manually extract goals from natural language documents [2]. The GBRAM has since been applied to financial and healthcare privacy policies [3]. Additional analysis of these extracted goals led to new semantics for modeling goals [8], [7], which distinguish rights and obligations, and new heuristics for extracting these artifacts from regulations [9], [10]. In regulations, a right describes an action that a stakeholder is permitted to perform, whereas an obligation describes an action they are required to perform. These distinctions are equivalent to permissions and obligations in Deontic Logic [15].

To facilitate reasoning with regulations, Antoniu et al. [4] introduced the regulations analysis method based on defeasible logic rules [20]. For this purpose, the facts manually found in the regulation document should be represented as a set of defeasible theory.

Given the need in facilitating the work of legal experts developing high quality legislations, standards and policies, a number of tools have been developed. We will briefly describe some of these tools. NormaSystem [21] is a user-friendly tool for creating and annotating legal documents. The system supports manual annotation activity and accepts input documents in such formats as are HTML, XML, RDF, and plain text. The tool then validates annotated documents according to the document structure (DTD and XML-schema validation), thus detecting inconsistencies in the semantic markup, for example, a missing publication date, duplication of the title or date, wrong content type. In a way similar to the user-friendly manner of annotation, the Norma-System provides the possibility to update a legislative document using the consolidation module. In turn, the Norma-Server serves not only as a repository for documents and metadata, but also manages versioning and provides some facilities for legal reasoning. The authors claim that reasoning module detects conditional modifications and uses *defeasible logic* to represent them.

MetaVex [23] is a regulation-drafting environment intended to be used by drafters and member of parliament. For this purpose, it provides editing facilities in a visual interface similar to a conventional word processor. The user starts creating the content in a word processor. In this stage, a set of templates structured according to the Dutch Guidelines for Legal Drafting can be used to facilitate the composing process. Elements that are frequently used in the domain of legal documents, such as citation, appendix, titles, and others, are factored out in a separate panel. Each of them can be instantiated by the user in an appropriate position. In addi-

tion, it provides the possibility for marking references to elements of (other) regulations and to individual entities, such as institutions or concepts defined by the regulation.

XMLegesEditor [1] is a legislative drafting environment developed to facilitate the adoption of Italian Legislative National XML Standards (NIR). The authors of the tool argue that existing WYSIWYG word-processors mainly focus on *style* markup rather than on structural and *semantic* markup. Therefore the original solution is proposed. In addition to providing a traditional word processor for creating the document content, the tool *a priori* guarantees generation of a valid XML document by constraining the user to perform only valid operations on the document. In order to support annotation using NIR elements, the tool provides a toolbar containing such elements.

The three editing systems considered above share many features. Each of them presents an original editing environment that allows legal experts to create and modify textual content in an understandable, transparent way. This means that users do not need to have any programming skills or knowledge of XML. XMLeges is principally oriented to developing documents according to the NIR standard. The modules of the system that automate semantic annotation of a legal document have been specifically trained to classify provisions and identify instances of the elements of according to this standard in texts written in the Italian language. On the other hand, Norma-System and MetaVex are not biased to a specific language. Though MetaVex provides a possibility to use the template based on the Dutch standards for legal writing to facilitate human's work, other templates can be incorporated.

To this end, the Cerno-based annotation process is complement to the regulation-drafting environments. Because Cerno allows constructing a generic process for analysis of documents, this feature makes the tool applicable to different types of regulatory texts, semantic models and languages. However, Cerno lacks a backend to the existing standards and, therefore, such environments, as for instance MetaVex and Norma-System, can serve as a useful tool for validation, storage and translation of the semantically annotated documents.

6. CONCLUSIONS

Regulations and policies constitute rich sources of requirements for software systems that must comply with these normative documents. In order to facilitate alignment of software system requirements and regulations, systematic methods and tools automating regulations analysis must be developed.

In the present work, we described the results of the application of Cerno, a lightweight framework for semantic annotation, to legal documents written in Italian. We investigated critical issues for semantic annotation tools in a different cultural and environmental context.

To evaluate the annotation results of the new process, we designed an empirical study, involving annotation of the Stanca law, and compared the performance of the tool with manual identification of instances of rights, obligations, and associated constraints. The results of this study are encouraging, and have also revealed a number of useful extensions for the tool and the tool-supported process. In summary, the proposed tool has demonstrated promising results with limited effort required to adapt it to a specific regulation document. Although, the phrase heuristics used are limited to the Stanca document and will need revision when analyzing other regulations and policies, we believe that our tool supported process can be re-used in a different domain due to its modularity. Further extensions and experimental evaluation are planned and being realized.

In particular, our future plans include extending the annotation framework to identify a wider range of concepts. We are also interested in developing reasoning facilities on the annotations using constraints of the domain meta-model, for instance, cardinality constraints. Finally, we will continue our experiments using different regulation documents.

Apart from the regulation compliance problem, another potential application of this work may be in providing support to lawmakers in writing high quality regulations in terms of improved consistency and reduced ambiguity. We believe that semi-automated tools such as the one proposed in this paper can be effectively used to improve the overall quality of rules and regulations at many levels.

Acknowledgments

7. REFERENCES

[1] T. Agnoloni, E. Francesconi, and P. Spinosa. XmLegesEditor: an OpenSource Visual XML Editor for supporting Legal National Standards. In *Proc. of V Legislative XML Workshop 2006*, European University Institute, Fiesole, 2007. Firenze, European Press Academic Publishing.

[2] A. I. Antón. Goal-based requirements analysis. In *Proc. 2nd IEEE Int. Conf. on Requirements Engineering*, pages 136–144, April 1996.

[3] A. I. Antón, J. B. Earp, Q. He, W. Stufflebeam, D. Bolchini, and C. Jensen. Financial privacy policies and the need for standardization. *IEEE Security and Privacy*, 2(2):36–45, 2004.

[4] G. Antoniou, D. Billington, and M. J. Maher. On the analysis of regulations using defeasible rules. In *Proc. 32nd Annual Hawaii Int. Conf. on System Sciences (HICSS'99)*, volume 6, pages 6033–6040. IEEE Computer Society, 1999.

[5] S. Berlusconi and L. Stanca. *Legge 9 gennaio 2004, n. 4: Disposizioni per favorire l'accesso dei soggetti disabili agli strumenti informatici*. CIAMPI, Rome, 2004.

[6] D. Berry, E. Kamsties, and M. M. Krieger. *From Contract Drafting to Software Specification: Linguistic Sources of Ambiguity–A Handbook*. Technical Report, School of Computer Science, University of Waterloo, Waterloo, ON, Canada, 2003.

[7] T. D. Breaux and A. I. Antón. Analyzing goal semantics for rights, permissions, and obligations. In *Proc. IEEE 13th Int. Requirements Engineering Conf.*, pages 177–186, Paris, France, 2005.

[8] T. D. Breaux and A. I. Antón. Deriving semantic models from privacy policies. In *Proc. 6th IEEE Int. Workshop on Policies for Distributed Systems and Networks (POLICY'05)*, pages 67–76, Washington, DC, USA, 2005. IEEE Computer Society.

[9] T. D. Breaux and A. I. Antón. Mining rule semantics to understand legislative compliance. In *Proc. of the 2005 ACM workshop on Privacy in the electronic society (WPES'05)*, pages 51–54, New York, NY, USA, 2005. ACM Press.

[10] T. D. Breaux, M. W. Vail, and A. I. Antón. Towards regulatory compliance: Extracting rights and obligations to align requirements with regulations. In *Proc. 14th IEEE Int. Requirements Engineering Conf.(RE'06)*, pages 46–55, Washington, DC, USA, 2006. IEEE Computer Society.

[11] J. Cleland-Huang, R. Settimi, X. Zou, and P. Solc. The detection and classification of non-functional requirements with application to early aspects. In *Proc. 14th IEEE Int. Requirements Engineering Conf.(RE'06)*, pages 36–45, Washington, DC, USA, 2006. IEEE Computer Society.

[12] T. R. Dean, J. R. Cordy, K. A. Schneider, and A. J. Malton. Using design recovery techniques to transform legacy systems. In *Proc. of IEEE Int. Conf. on Software Maintenance (ICSM 2001)*, pages 622–631, November 2001.

[13] P. del Consiglio dei Ministri. Guida alla redazione dei testi normativi. *Gazzetta Ufficiale*, 101(2):105, 2001.

[14] U. S. Goverment. Standards for privacy of individually identifiable health information, 45 CFR part 160, Part 164 subpart E. *In Federal Register*, 68(34):8334–8381, February 20 2003.

[15] J. F. Horty. *Agency and Deontic Logic*. Oxford University Press, New York, NY, 2001.

[16] N. Kiyavitskaya, N. Zeni, T. D. Breaux, A. I. Antón, J. R. Cordy, L. Mich, and J. Mylopoulos. Extracting rights and obligations from regulations: Toward a tool-supported process. In *Proc. of 22nd Int. Conf. on Automated Software Engineering*, pages 429–432, 2007.

[17] N. Kiyavitskaya, N. Zeni, L. Mich, J. R. Cordy, and J. Mylopoulos. Applying software analysis technology to lightweight semantic markup of document text. In *Proc. of Int. Conf. on Advances in Pattern Recognition (ICAPR 2005)*, pages 590–600, 2005.

[18] N. Kiyavitskaya, N. Zeni, L. Mich, J. R. Cordy, and J. Mylopoulos. Text mining through semi automatic semantic annotation. In *Proc. of Practical Aspects of Knowledge Management (PAKM'06)*, volume 4333 of *LNCS*, pages 143–154. Springer-Verlag, 2006.

[19] B. Moulin and D. Rousseau. Knowledge acquisition from prescriptive texts. In *Proc. 3rd Int. Conf. on Industrial and engineering applications of artificial intelligence and expert systems (IEA/AIE '90)*, pages 1112–1121, New York, NY, USA, 1990. ACM Press.

[20] D. Nute. Defeasible reasoning. In *Proc. 20th Hawaii Int. Conf. on Systems Science*, pages 470–477. IEEE Press, 1987.

[21] M. Palmirani and R. Brighi. Norma-System: A Legal Document System for Managing Consolidated Acts. In *Proc. of the 13th Int. Conf. DEXA 2002*, pages 295–314, September 2-6 2002.

[22] H. Schmid. Probabilistic part-of-speech tagging using decision trees. In *Proc. of Int. Conf. on New Methods in Language Processing*, Manchester, UK, 1994.

[23] S. van de Ven, R. Hoekstra, and R. Winkels. MetaVex: Regulation Drafting meets the Semantic Web. In *Proc. of the Workshop on Semantic Web technology for Law (SW4Law 2007)*.