An Approach Based on Citation Analysis to Support Effective Handling of Regulatory Compliance

Mohammad Hamdaqa and Abdelwahab Hamou-Lhadj
Department of Electrical and Computer Engineering
Concordia University
1455 de Maisonneuve West Blvd.
Montréal, QC, Canada
{m_hamdaq, abdelw}@ece.concordia.ca

Abstract

For most global software companies with a client base that covers a large number of regulated businesses, regulatory compliance represents a significant challenge. The world of compliance has become increasingly complex due to the overwhelming number of regulations, laws, and standards that are introduced every year. These laws may vary significantly in their scope and applicability depending on the industry sector and the geographical area of the end client. In addition, many of these laws are created by different legislative bodies resulting in overlapping and sometimes conflicting provisions. To further complicate matters, laws are often created based on existing ones, forming a complex set of interdependent rules where changes made in one place can propagate to affect, sometimes in an inconsistent manner, many other laws. There is clearly a need to investigate techniques and tools that can alleviate IT solution providers from the complexity of dealing with regulatory compliance. In this paper, we present an approach and a supporting tool that aim to facilitate the analysis of multiple regulations. Our approach is based on the exploration of the citation relationship that links various laws together. The citation relationship is represented by a citation graph that can be used by an analyst to navigate through the provisions of various interrelated laws to uncover overlaps and possible conflicts or to simply understand the content of specific law documents. We also present a tool called CompDSS (Compliance Decision Support System) that supports our approach. Finally, we show the effectiveness of the presented approach by applying it to three regulations, namely, SOX, HIPAA, and GLBA.

Keywords:

Software Engineering; Regulatory compliance; IT Compliance; Citation Analysis.

1. Introduction

For many regulatory companies, regulatory compliance has become an important part of their business regardless of geography and industry sector. There are a number of factors behind the recent increase of attention to regulatory compliance including corporate scandals and the need for accountability, the removal of trade barriers, the reliance on Information Technology (IT), the necessity to protect and secure sensitive information [Oppliger 2000, Chang 2004], and a higher need for business continuity and assurance [Silverman 2008, Ernst 2006]. As a result, more and more authoritative rules (i.e., regulations, laws, standards, and guidelines) are introduced every year putting further constraints on the way companies are operated, managed, controlled, and governed [Hamou-Lhadj 2007]. The consequences of not complying with these laws can be devastating and may include substantial fines, financial losses, lawsuits, customer dissatisfaction, and loss of reputation and market confidence. Examples of some of the most popular North-American laws include the Health Insurance Portability and Accountability Act (HIPAA), the Personal Information Protection and Electronic Documents Act (PIPEDA), the Gramm-Leach-Bliley Act (GLBA), the Sarbanes-Oxley Act (SOX), the California Data Protection Act, and many others.

Many of these regulations and laws have a direct an impact on the way software systems used by regulated companies are developed, tested, and maintained. For example, a data records management tool, used by a health institution which is required to comply with HIPAA, must support data security features such as authentication mechanisms, different levels of data access control, and frequent backups and data reliability techniques. A recent survey of IT tools with a focus on privacy and security shows that regulatroy compliance is one of the key requirements that need to be taken into account when building such tools [Ouellet 2009].

Cascading the compliance needs to software companies has put technology builders in a new business context in which they are required to develop software products that meet compliance requirements while reducing the cost of development and maintenance. For most global software companies, this represents a significant challenge. First, there are just too many laws to comply with, especially that a typical global software company can have hundreds (if not thousands) of clients worldwide spread over several heavily regulated industry sectors. In his description of the U.S. regulatory compliance landscape in the area of data privacy and security alone, Silverman notes that "Given the enormous breadth of federal regulation, it is not possible to catalogue the full range of U.S. law pertaining to data, privacy, and records" [Silverman 2008]. He adds that the issue is further complicated considering that each state has its own set of regulations and that many of these laws are often developed by different legislative bodies with little effort for consistency or convergence with similar legal requirements, which often results in duplicated and conflicting rules. The situation is generalized to many other areas of law and it is made even worse for those companies that operate at the global level and which must also comply with international regulations. In addition, regulations are rarely created from scratch. They often refer to other existing laws, creating a complex web of interrelated rules in which changes made in one place can have a marked effect elsewhere. For example, SOX, GLBA, and HIPAA refer to many common laws including the Security Exchange Act, the Securities Investor Protection Act, the Labor Act, and the Employee Retirement Income Security Act. Amendments made by SOX, for instance, to any of these common laws will most likely affect GLBA and HIPAA. In such a situation, any corporate policy, tool feature, or business process based on GLBA and HIPAA will need to be reviewed along with the changes caused by SOX.

Clearly, there is a need to investigate ways to help IT companies manage a large number of possibly overlapping or conflicting regulatory compliance requirements. This translates into the need to represent and organize regulatory compliance documents in such a way that it is easier to explore and analyze so as to enable users (e.g. software engineers, software process engineers, project managers, etc.) extract and prioritize the main provisions, uncover similarities and conflicts among inter-related regulatory compliance documents, check for consistency due to law amendments, etc. These are the objectives of our research, which pertains to an emerging field of study that is referred to as Software Compliance Engineering[1] which is defined as the software engineering field that is concerned with investigating techniques and tools to develop, test, and maintain software systems where compliance is a built-in quality attribute.

[1] http://users.encs.concordia.ca/~abdelw/softwarecompliance.html

In this paper, we present our latest contribution to the field of Software Compliance Engineering which consists of an approach and a supporting tool that facilitate the analysis of multiple regulations. Our approach is based on examining the relationship between law documents by exploring the citation relationship that connect them. The work presented in this paper is a continuation of our previous work, published as a position paper [Hamdaqa 2009], where we discussed our thoughts on the importance of analyzing the citations in law documents in order to facilitate the understanding of their content. In this paper, we show how citation analysis can be a powerful tool for detecting overlaps and possible conflicts among laws, as well as assessing the impact of changes made by some laws on the other ones that relate to it. Another contribution of this paper is a detailed approach for extracting citation graphs from multiple regulations and which involves extracting the provisions, the citations, and the types of relations among them. In addition, we present a prototype tool, called CompDSS (Compliance Decision Support System), which supports citation analysis as one of its main features. A final contribution of the paper is a case study in which we show the effectiveness of our approach to detect conflicts and overlaps in three regulatory documents, namely, SOX, HIPAA, and GLBA.

There are various types of legal documents including cases, statutes, and administrative regulations. Cases are based on a judicial decision that is reported in the countries that use common law systems. Statutes are enacted by legislature and used to grant authority to administrative agencies to adopt and amend administrative. The focus of this research is on statutes and administrative laws (that we refer to interchangeably as laws and regulations). Also in this paper, we limit ourselves to North-American laws and regulations. However, we believe that the proposed approach can be easily adapted to other laws.

Organization of the paper: In the next section, we define the concepts of citation and citation analysis. We also present a detailed approach for extracting citation graphs from multiple documents. The section ends with a process that describes how to apply citation analysis to detect conflicts and overlaps. In Section 3, we present the tool and discuss its main components. The case study is presented in Section 4, followed by related work in Section 5. We conclude the paper and discuss the limitations of our approach and how to address them in Section 6.

2. Citations in Legal Documents

2.1. What is a Citation?

A citation describes a relationship between two documents (or parts of these documents), where one document (the citing) refers to another document (the cited). Legal citations are citations found in legal documents, which usually connect the provisions of one document to the provisions of either the same document (internal citations) or a different one (external citations). Legal citations play an important role in enforcing the legitimacy of the arguments and propositions contained in legal documents while reducing the space needed to write these documents [AALL 2002].

Proper citation of legal documents has always been a difficult task [AALL 2002]. To help with this process, a number of citation manuals containing a set of comprehensive rules have emerged, among which the most popular ones in the U.S. are perhaps the Bluebook [Barris 2007] and the ALWD (Association of Legal Writing Directors) [Dickerson 2003] manuals. In Canada, the most common citation manual is the Canadian Guide to Uniform Legal Citation published by McGill Law Journal [CGULC 2006].

An example of a citation that follows the Bluebook standard is "15 U.S.C. § 78u(d)(3)(B)(iii)(II)(aa)". This citation is taken from the Sarbanes-Oxley Act and refers to the U.S. Code, which is a repository of U.S. laws identified with unique codes. In this example, "15 U.S.C. § 78u" refers to Title 15 of the Commerce and Trade Act, which consists of the Securities Investor Protection Act Section 78. A citation consists of three main components: The volume or title number, the name of the regulation, and the section number, which is usually preceded by the section sign "§". The section can be further divided into subsections, paragraphs, subparagraphs, clauses, and sub-clauses.

2.2. Citation Analysis

We define citation analysis as a technique that enables the examination of the relationship among regulatory compliance documents through the exploration of the citation relationship that connect their respective provisions. Citation analysis relies on a citation graph, extracted by parsing regulatory documents. The graph nodes represent the provisions and the edges represent the citing relationship. There are many advantages in exploring the relationships among provisions within the same document or across multiple documents including:

- Assessing the impact of a change in a particular act on which many other acts depend on. A citation graph, supported by a usable tool, can readily be used to understand the possible impact of these changes by highlighting the places in the different acts that are potentially impacted.
- Citation analysis can be used to check the consistency of multiple acts that make different modifications to a common act on which they depend. For example, if two acts A and B depend on a third act C and that A modifies C's provisions then B might be affected by the changes made by A. From our experience, most conflicts that exist in regulations are due to lack of tools that assist users in doing consistency checks. We believe that citation analysis, if supported effectively by a tool, can be a powerful solution to this problem.
- Detecting overlaps and possible conflicts that exist between multiple regulatory documents through analyzing the provisions (and the respective acts) they have in common.
- Finally, citation analysis can help understand regulations by allowing easy navigation through its provisions and the ones from other acts on which it depends on. To achieve this, a tool that supports citation analysis should allow for features such as searching and slicing based on internal and external citations, limiting the depth of a citation (e.g., section, subsection, paragraphs, etc.), and so on.

2.3. Building a Citation Graph

A citation graph is a directed non-ordered graph G = (V, E, R) where:

- V: Represents a set of nodes which represent the citing and the cited provisions.
- E: Represents a set of edges. An edge between Node A and Node B exists if A has a citation to B.
- R: Represents a set of relations between two nodes. There are various types of relations among provisions such as amendments and assertions. These relations are discussed in Section 2.3.2 when the proposed algorithm for extracting relations is described. Relations appear as labels on the graph's edges.

2.3.1. *Extracting the Provisions*

Our definition of a provision is similar to the definition provided by Eugen Ehrlich in his classic paper “The Sociology of Law” [Ehrlich 1922], where he defines a provision as “an instruction framed in words addressed to courts as to decide legal cases or a similar instruction addressed to administrative officials as how to deal with particular cases” [Ehrlich 1922]. A provision can be in the form of a clause, sentence, or paragraph of a legal document, which provides information for a particular matter. Provisions are considered the core of regulations since they contain the rights and obligations, assets, and liabilities stipulated in the corresponding act. Citations are embedded in the provisions as one of their main components, and play a critical role in understanding the provisions [Hamdaqa 2009].

Our process of extracting provisions from regulatory documents is based on the fact that most North-American legal documents use indentation to indicate the beginning of a section, its subsections, paragraphs, etc, forming a paragraph hierarchy structure. However, since most documents are usually saved in an unstructured manner (mostly in PDF and HTML) and come without a table of content, we needed to create a set of regular expressions that can identify the various components of a provision during the parsing of the law document under study. These regular expressions are shown in Figure 1. Each provision is uniquely indexed using a combination of the document title, section, subsection, and other subcomponents.

```
TITLE I: TITLE\s?[IVX]{1, 3}
   SEC.: SEC.\s?\d{1,5}
         (a): \([a-hj_z]\)
            (1): \(\d{1,2}\)
               (A): \(A-HJ-Z)\)
                   (iii): \([ivx]{1,3}\)
                       \([IVX]{1,3}\)
                           (aa): \([a-h][a-h]\)
```

Figure 1. The regular expressions used to extract provisions according to the paragraph hierarchy structure

The process of extracting provisions using the set of the regular expressions includes searching the regulation document for the patterns defined in Figure 1, which we have developped based on the United States Congress Data Dictionary of Legislative Documents[2]. For example TITLE\s?[IVX]{1,3} represents the word TITEL followed by an optional space and then a roman number that ranges from one digit to a maximum of three. The expression \(\d{1,2}\) is equal to one or two digits between two brackets which represents a subsection according to the paragraph hierarchy structure. After identifying all parts of the document, we organize the document in a hierarchical tree structure to simplify its processing. Each provision will be indexed using its full path from the title down to subparts in order to identify the location of the information in the same way it is cited in other legal documents. For example the provision identifier "Title I.Sec.102(a)" of SOX can help the reader to navigate through the document to reach this particular provision.

2.3.2. Extracting the Citations

The next step is to extract the citations. In this paper, we limit ourselves to the ones that follow standardized citation styles such as the Bluebook and ALWD. For this purpose, we have carefully developed a set of regular expressions that capture the structure of citations expressed in these two standards. The resulting regular expression is represented in the automata of Figure 2.

[2] http://xml.house.gov/

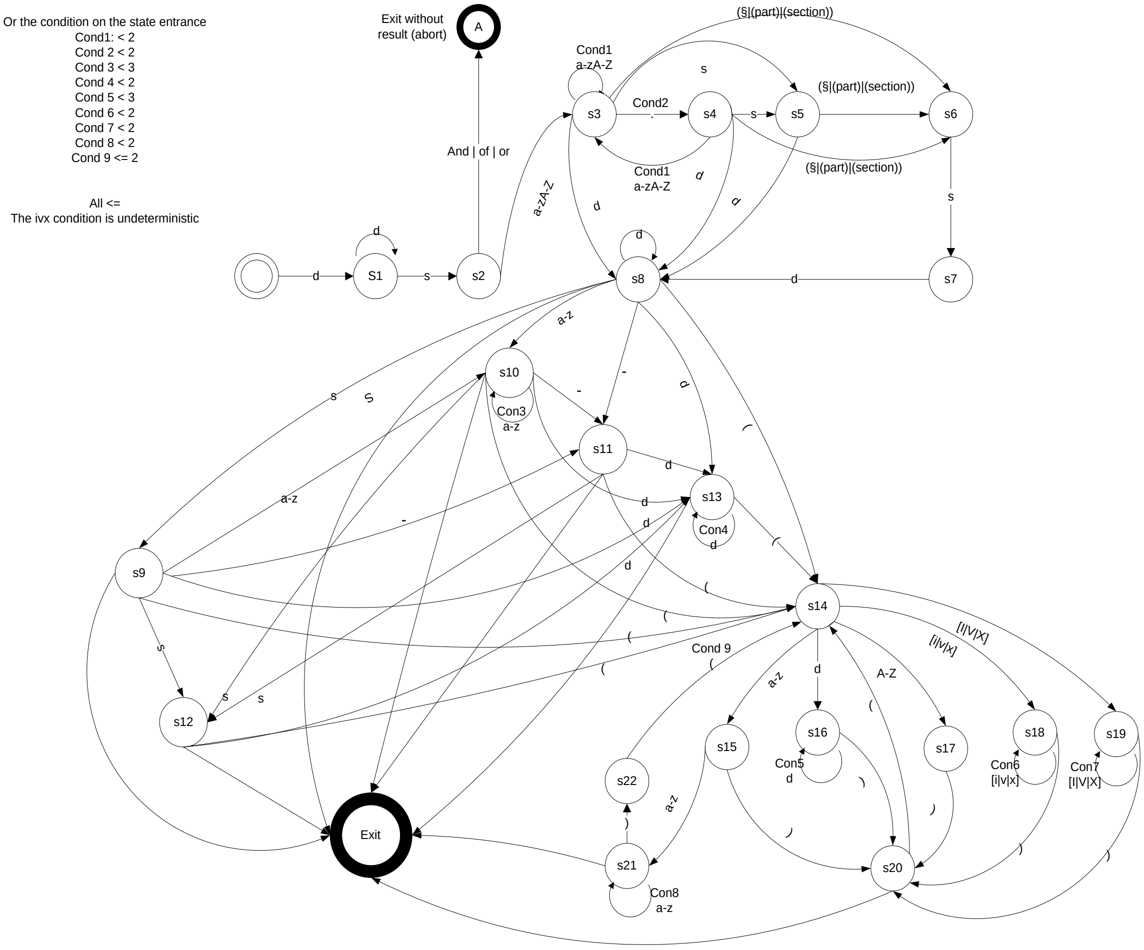


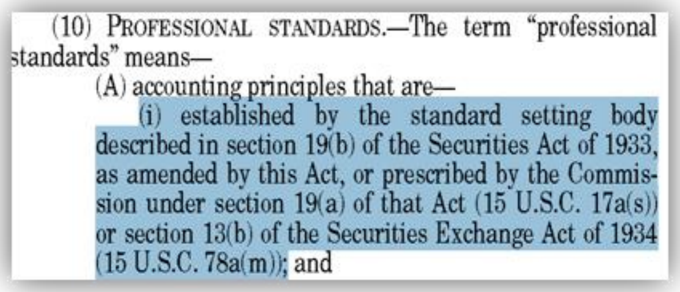
Figure 2. Automata for the regular expression that captures the standardized citation styles

The automata in Figure 2 is built to recognize citations written according to the Bluebook and ALWD standardized citation styles. We carefully studied these standards to include all possible combinations that may be recognized as a citation. For example, a typical citation that follows these standards (e.g. the citation "5 U.S.C. § 78u(d)(3)(B)(iii)(II)(aa)") starts with a number followed by a space, followed by a maximum of three characters (which is an abbreviation of the law to which the citation refers to), a space, a '§' sign, and the parts that form the section, subsections, etc. To ensure that we capture only the abbreviations permitted by the Bluebook and ALWD standard, we used the GovSpeak[3] library which is a widely accepted resource that maintains most abbreviations and acronyms of the U.S. Government.

[3] http://members.cox.net/govdocs/govspeak.html

It should be noted that, after studying many regulatory documents, we have realized that the standards have not always been followed. For example, in the SOX act, the section sign (§) is not used consistently. We therefore needed to create a regular expression that accounts for citations that have a structure that is similar but not necessarily identical to the one provided by the Bluebook and ALWD standardized citation styles.

Another source of ambiguity when dealing with citations is the use of what is known as parallel citations, which occurs when the same provision contains two consecutive citations that refer to the exact same information found in two other different documents. Although not frequent, parallel citations appear in the context of court cases, where the first cited document is the original and official source of the information, whereas the second one refers in many situations to a document archived by a private company to enforce the existence of the cited provision.



(10) PROFESSIONAL STANDARDS.—The term "professional standards" means—

(A) accounting principles that are—

(i) established by the standard setting body described in section 19(b) of the Securities Act of 1933, as amended by this Act, or prescribed by the Commission under section 19(a) of that Act (15 U.S.C. 17a(s)) or section 13(b) of the Securities Exchange Act of 1934 (15 U.S.C. 78a(m)); and

Figure 3. Example of a parallel citation found in SOX

The example in Figure 3 shows a parallel citation, where both cited documents, the "Securities Exchange Act of 1934 section 13(b)" and the "15 U.S.C. 78a(m)" of the United States Codes, contain the exact same information. In this paper, we only focus on citations that follow the standard citation styles. The detection of citations that do not follow standards is left as future work.

2.3.2. Extracting the Relations

The next step is to determine the relations between the citing and the cited provisions. As mentioned earlier, there exist many types of relations that link a citing provision to a cited one. After manually inspecting many regulations, we found that these relations can be grouped in two main categories: Assertions and Amendments. A relation is considered as an assertion if it refers to a provision that is used to support the author’s point of view through examples, definitions, or any other additional information. An assertion relation can be further divided into the following subtypes:

- Definition: This is the case where the cited provision defines the citing provision.
- Specification: This is the case where the cited provision provides more information about the citing provision.
- Compliance: This relation indicates a cited provision that complies with the citing one.

A citation is considered as an amendment if the citing provision amends the cited provision (or part of it). Amendments can be further divided into subtypes including:

- Amendment by insertion: The citing provision adds more details or complete parts to the cited provision.
- Amendment by deletion: The citing provision deletes parts of the cited provision.
- Amendment by striking: This relation is used to attract the reader’s attention by crossing the information about a cited provision that is no longer valid and inserting new parts or details.
- Amendment by redesignation: This occurs when the cited provision changes the name. The new name is then reflected in the citing provision.

It is worth noting that amendments often appear more frequently than assertions. This is due to the fact that they are used as a legal instrument for changing the content of a particular act to adapt it to changing circumstances. Suber notes that “Amendments aspire to capture the inconsistent virtues of stability and flexibility, protecting what the enacting generation thinks wise, but permitting future generations to think otherwise” [Suber 1999]. An act can amend itself or other acts. In this study, we will pay careful attention to amendments since they are the ones that can potentially generate overlaps and conflicts [Antoniou 1999].

We can identify the relation that a citation represents by exploring the text surrounding it searching for specific verbs such as “amends”, “defines”, etc. To extract these verbs, we use a text tagging technique based on Brill's Transformation-based learning tagger to perform the Part of Speech (POS) tagging process [Brill 1995]. The objective is to divide the text into words, and assign each word its corresponding part of speech (i.e., noun, verbs, preposition, etc.), based on both its definition, as well as its context. These words can later be explored individually for the purpose of extracting the type of relations among provisions. An example of tagging the provision of Figure 4a is shown in Figure 4b.

TITLE XI SEC.1103(b)

TECHNICAL AMENDMENT.—Section 21C(c)(2) of the Securities Exchange Act of 1934 (15 U.S.C. 78u–3(c)(2)) is amended by striking ‘‘This’’ and inserting ‘‘paragraph (1)’’

TITLE XI SEC.1103(b)

TECHNICAL/NNP AMENDMENT.—Section/NNP 21C(c)(2)/CD of/IN the/DT Securities/NNP Exchange/NNP Act/NNP of/IN 1934/CD (*) is/VBZ amended/VBN by/IN striking/JJ ‘‘This’’ and/CC inserting/VBG ‘‘paragraph (1)’’

Figure 4. a) Example of a provision taken from SOX before tagging b) Example a tagged provision

Before the tagging process starts, we replace all citations by a (*) symbol so they can be treated as one word. The citations are put back in the document after the tagging process is complete. After that, we collect the verbs with direct relations to the citation of interest. Using the WordNet [Miller 1995] lexical database cognitive synonyms, which is a set of one or more synonyms that can be used interchangeably without changing the meaning of the context in which it is embedded, we were able to convert these verbs to one of the relations we discussed earlier (i.e., assertion or amendment). The morphological processor in WordNet finds the base form automatically and then returns all synonyms of the word. The last step is to match the meaning of the verb with one of the relations we already defined. An example of using WorldNet synonyms is shown in Figure 5, in which the synonyms of the word “Revise” (base form of “Revised”) are listed, among which the word “amend” appears, which would classify the citation around which the word “revised” appears as an amendment

```
Word    ➔ Base   ---- (POS) synonym1 | synonym2

Revised ➔ Revise ---- (n) revise | rescript | revisal | revision|
                 ---- (v) rewrite | retool | amend
```

Figure 5. Example of the matching process used in extracting the relations

2.3.3. Graph Building

The final step of building the citation graph is to generate the graph itself. To generate the graph we need the elements generated in the previous steps, which are the citing and the cited provisions, and the relations between them. We use the DOT scripting language [Graphviz] to describe the graph. The DOT language allows expressing the nodes, edges, and labels that will form the final graph (see Figure 7a). Graphviz supports two graph rendering techniques the circular graph layouts “circo” and the spring model layout “neato”. These techniques differ in he final layout and the efficiency of the rendering process. It is recommended to use the “neato” approach when the number of edges of the graph is relatively high for a large number of nodes. Figure 7a and 7b show the DOT file and the final graph that correspond to the citation graph constructed from the SOX provisions shown in Figure 6.

SEC. 2. DEFINITIONS

(a) IN GENERAL …

(10) PROFESSIONAL STANDARDS.—The term ‘‘professional standards’’ means—

(A) accounting principles that are—

(i) established by the standard setting body   described in section 19(b) of the Securities Act of 1933, as amended by this Act, or prescribed by the Commission under section 19(a) of that Act (15 U.S.C. 17a(s)) or section 13(b) of the Securities Exchange Act of 1934 (15 U.S.C. 78a(m));

SEC. 108. ACCOUNTING STANDARDS.

(a) AMENDMENT TO SECURITIES ACT OF 1933.—Section 19 of the Securities Act of 1933 (15 U.S.C. 77s) is amended—

(1) by redesignating subsections (b) and (c) as   subsections (c) and (d), respectively; and

(2) by inserting after subsection (a) the following: ‘‘(b) RECOGNITION OF ACCOUNTING STANDARDS.—

Figure 6. Example of two provisions taken from SOX

```
digraph CitationGraph
{
    {node [style=filled,color= skyblue] "TITLE I SEC.108(a)(2)"}
    "TITLE I SEC.108(a)(2)" -> "15 U.S.C. 78a(m)" [label= "Amend By Redesignation"];

    {node [style=filled,color= skyblue] "TITLE I SEC.2(a)(10)(A)(i)"}
    "TITLE I SEC.2(a)(10)(A)(i)" -> "15 U.S.C. 17a(s)" [label= "Described by"];
    "TITLE I SEC.2(a)(10)(A)(i)" -> "15 U.S.C. 78a(m)" [label= "Described by"];
    "TITLE I SEC.2(a)(10)(A)(i)" -> "SA of 1933 Section 19(b)" [label= "Described by"];

    {node [style=filled,color= skyblue] "TITLE I SEC.108(a)(1)"}
    "TITLE I SEC.108(a)(1)" -> "SA of 1933 Section 19(b)" [label= "Amend By Insertion"];
}
```

Figure 7a. A citation graph represented in DOT language extracted from the provisions described in Figure 6

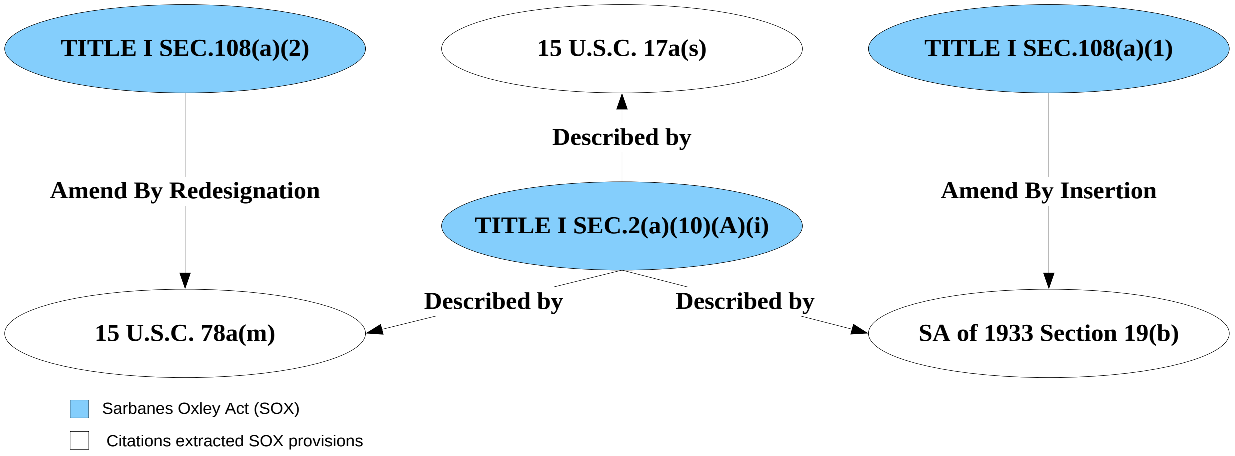


Figure 7b. A citation graph extracted from the provisions described in Figure 6

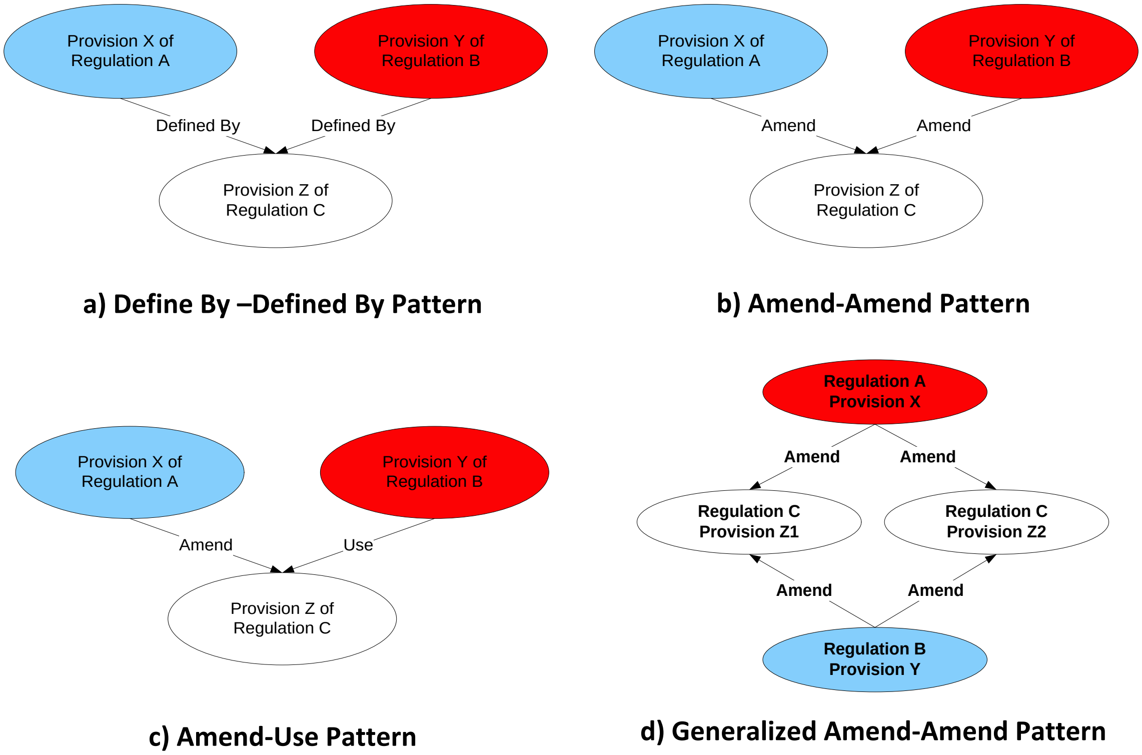
2.4. The Process of Applying Citation Analysis

We propose a top-down process for applying citation analysis to detect how different acts relate to each other, starting by exploring the citation graph at a high-level looking for places where regulations are intersect (e.g. by citing common laws), and digging into a more detailed graph that shows the exact sections and the type of relations that connect the citing and the cited regulations. The process can be summarized in the following steps:

1. We analyze the citation graph to identify regulations that are related to each other through the citation relationship.
2. For each of the intersecting regulations, we dig one level deeper to study the provisions that relate to each other. This is because two citing provisions may refer to the same act, but to different provisions in the cited act.
3. For each related provisions, we analyze whether it is an amendment or an assertion. As previously mentioned, amendments, in particular amendments by deletion and striking, represent high risks of conflict since an act that deletes parts of another act will affect all other acts that depend on the modified act.

After studying several regulatory documents and the dependencies among them, we have defined a set of graph patterns that represent situations where possible overlaps and conflicts may occur. These patterns take into account the related provisions and the type of relations that connect them (i.e. assertions and amendments). They constitute by no means an exhaustive list of patterns but are, in our view, very representative to the cases we encountered in practice. It should also be noted that our approach does not aim to detect automatically all conflicts and overlap, instead, it should be used to guide the analysts when exploring law documents by indicating places where conflicts and overlaps may occur. It is therefore important that a tool that supports these patterns, however, needs to allow enough flexibility to the user to add new ones. The patterns are presented in what follows.

- **The Defined By–Defined By Pattern:** When two provisions X and Y from two different regulations A and B are respectively using the definition or the concept explained in another regulation C (see Figure 8a), this indicates a possible similarity between provision X and provision Y. As a result, there is high possibility to find common rules between the two provisions and as such any changes to A needs to be reflected in B and reciprocally.
- **The Amend-Amend Pattern**: When two provisions X and Y from two different regulations A and B respectively amend a provision from another regulation C (Figure 8b), this may result in a conflict, if they modify it in a contradictory way. Amendment by redesignation does not affect the content so it should not be considered when looking for conflicts. However; amendment by striking, which is usually followed by an amendment by insertion or deletion, can lead to a conflict since they change the content of the amended law.
- **The Amend-Use Pattern:** If a provision X from an act A amends a provision Z from C and that a provision Y of act B uses the provision C then changes made by A may affect B (Figure 8c). This may result in inconsistencies and conflicts.
- **The Generalized Amend-Amend Pattern:** We have noticed through the study of many regulations that related provisions from different acts usually refer to the same set of provisions. In other words, if a provision Y from B refers to provision Z1 of regulation C, whenever the provision X of regulation A refers to Z1, then it is most likely that provision X will also refer to provision Z2 of regulation A when provision Y refers to Z2 (Figure 8d). This is a more generalized case of the Amend-Amend pattern.


Provision X of Regulation A
Provision Y of Regulation B
Defined By
Defined By
Provision Z of Regulation C
Provision X of Regulation A
Provision Y of Regulation B
Amend
Amend
Provision Z of Regulation C
Regulation A Provision X
Amend
Amend
Provision X of Regulation A
Provision Y of Regulation B
Regulation C Provision Z1
Regulation C Provision Z2
Amend
Use
Amend
Amend
Provision Z of Regulation C
Regulation B Provision Y

a) Define By –Defined By Pattern

b) Amend-Amend Pattern

c) Amend-Use Pattern

d) Generalized Amend-Amend Pattern

Figure 8. Citation analysis patterns for detecting similarities and conflicts

In what follows, we show how these patterns can be effective in detecting conflicts among regulations by applying to detect a well-documented and controversial conflict in the area of law between two the copyright laws, The Digital Millennium Copyright Act (DMCA) of 1998, and the Technology Education and Copyright Harmonization Act (TEACHA) of 2002. The problem was first reported in 2003 by a group representing the college media centers, which sent a warning message to the U.S. Copyright Office about a possible conflict between these two federal laws.

The conflict consists of the fact that while DMCA restricts access to the electronic copyrighted material, TEACHA allows access to the same material under the “fair use” cover for pedagogical purposes, more specifically for online education and distance learning. The “fair use” umbrella aimed to solve the problem if TEACHA did not restrict the conversion of materials from analog to digital format, which will be considered a violation of the anti-circumvention provision which was added to the copyright act by the DMCA.

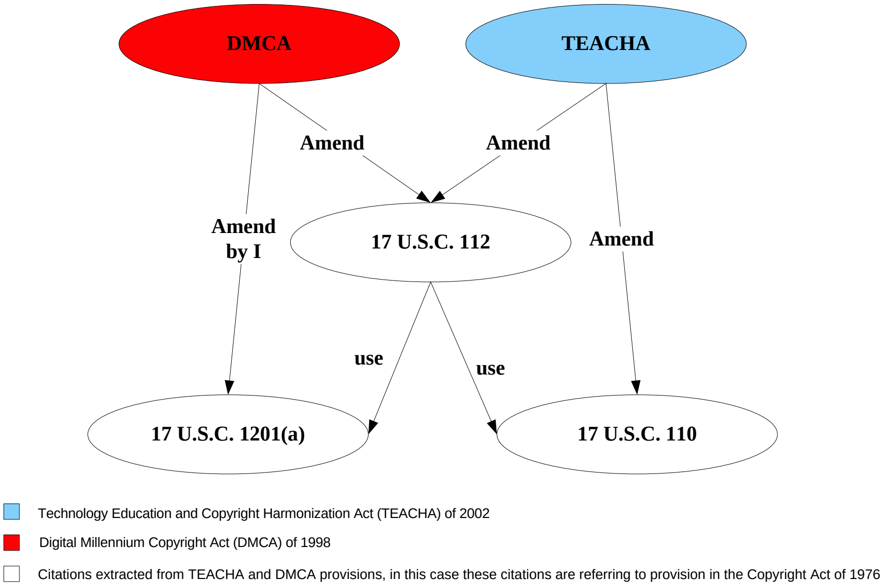

DMCA
TEACHA
Amend
Amend
Amend
by I
17 U.S.C. 112
Amend
use
use
17 U.S.C. 1201(a)
17 U.S.C. 110
🟦 Technology Education and Copyright Harmonization Act (TEACHA) of 2002
🟥 Digital Millennium Copyright Act (DMCA) of 1998
□ Citations extracted from TEACHA and DMCA provisions, in this case these citations are referring to provision in the Copyright Act of 1976

Figure 9. The citation graph that shows a situation where a possible conflict between DMCA and TEACHA may exist

We applied citations relatedness analysis on DMCA and TEACHA to uncover this conflict. Figure 9 shows a partial graph representing the provisions related to access to the electronic copyrighted material in both laws. The graph combines the first three patterns explained earlier. Using this graph, we can see that TEACHA amended Section 110 as well as Section 112 of the Copyright Act Title 17 of the U.S. code, at the same time the DMCA amended Section 112 and added a new section, Section 1201 to the Copyright Act. Based on citation relatedness analysis, this situation represents a high risk of non-compliance due to the dependencies between Sections 112, 110 and 1201. Further analysis is needed to see if this dependency yields a conflicting situation.

According to the graph the Copyright Act of 1976 Title 17 of the U.S. code was amended by both DMCA and TEACHA. The Copyright Act is an old act; hence it does not cover the issues brought up by the new technologies such as protection of digital content. The objective for TEACHA amendment of Section 110 (Limitations on exclusive rights: Exemption of certain performances and displays) is to allow instructors who work in non-profit educational institutions to use portions of movies, or audiovisual work in online courses without looking for permission or paying fees to copyright holders. The specified amendment clarifies the conditions under which such usage will not be considered as infringement of the copyright.

DMCA added Section 1201(a) which is referred to as the anti-circumvention provision. The provision makes it illegal to circumvent technologies that prevent access to copyrighted material such as ripping Macrovision or Content Scramble System DVDs. On the other hand the DMCA released some of the liabilities of transmitting organizations and added a statutory license subsection to Section 112.

TECHA allows using the copyrighted materials in distance learning, as a result there is a need to have the copyrighted materials in digital format. DMCA restricts the circumvention of technologies and converting materials from analog to digital format. There is a need for some exceptions for distance learning so that TEACHA can achieve its goals. The amendment by TEACHA to Section 112 - Limitations on exclusive rights: Ephemeral recordings - restricted the conversion of materials from analog to digital format unless the following conditions hold.

> *"(A) no digital version of the work is available to the institution; or*
>
> *(B) the digital version of the work that is available to the institution is subject to technological protection measures that prevent its use for section 110(2)."*[4]

These conditions complicate the problem especially that not all digital material that is protected has a non-digital version that educational institutions can convert to make use of TEACHA. The contradictions between the two copyright laws did not provide the instructor's with a safe path while practicing distance learning, and kept the liability of their actions up to the judges' decision based on the "fair use" act.

3. Tool Support: CompDSS

3.1 Overview of CompDSS

We have developed a prototype tool, called CompDSS (Compliance Decisions Support System), which supports the techniques investigated by our research group. Using CompDSS, a user can load multiple regulations for exploration, display the list of provisions they contain, generate citation graphs, and investigate the relationships among regulations. The tool also supports standard features such as search capabilities and multiple exploration views.

[4] Subparagraph (A) and (B) of Subsection (f) of Section 112 of title 17, United States Code as amended by TEACHA

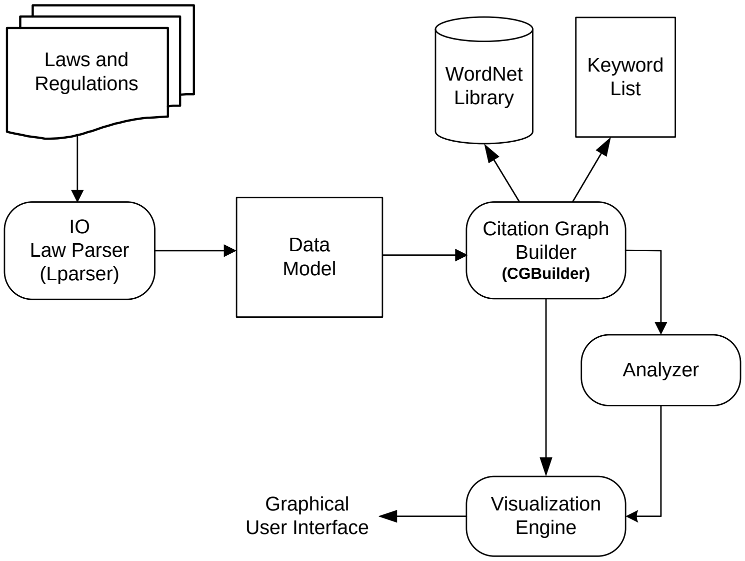


Figure 10. Overall Architecture of CompDSS components that support citation analysis

The overall flow of information among the CompDSS components that support citation analysis are shown in Figure 10, and consists of a law parser (LParser), a citation graph builder (CGBuilder), an analyzer, and a visualization engine.

The role of the parser is to extract the provisions and their constituents from an input law document and build a data model that can later be processed by the other components. This requires some preprocessing steps to clean up the input documents, which are usually saved in unstructured formats such as PDF and XML as mentioned earlier, by removing headers, footers and other unnecessary data. The parser relies on the regular expression discussed in Section 2.3.1 and which is designed specifically to recognize the various elements of a North-American law. The user has, however, the flexibility to modify these regular expressions to adapt them to other laws.

The output of the parser is an object model that instantiates a simple and yet expressive data model that we built to characterize the content of law documents. The data model is shown in Figure 11 in the form of a UML class diagram. In this model, an act is composed of several provisions, which contain various elements organized in a hierarchical way including sections, subsections, clauses, sub-clauses, and so on. This hierarchical structure is captured using the composite design pattern [Gamma 1994] and is modeled using three classes, namely, ProvisionElement (abstract class), LeafElement, and CompositeElement.

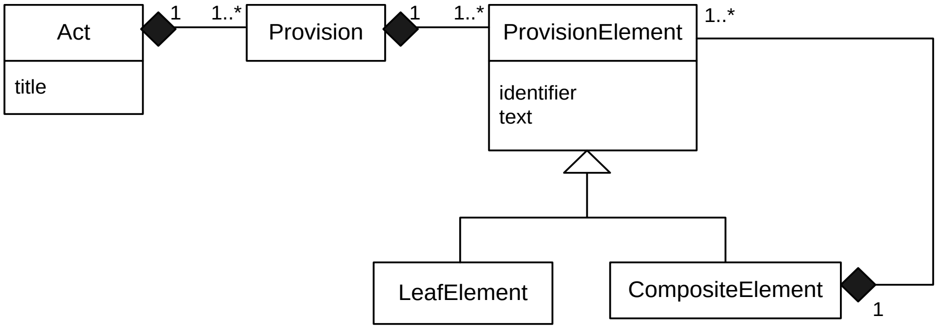
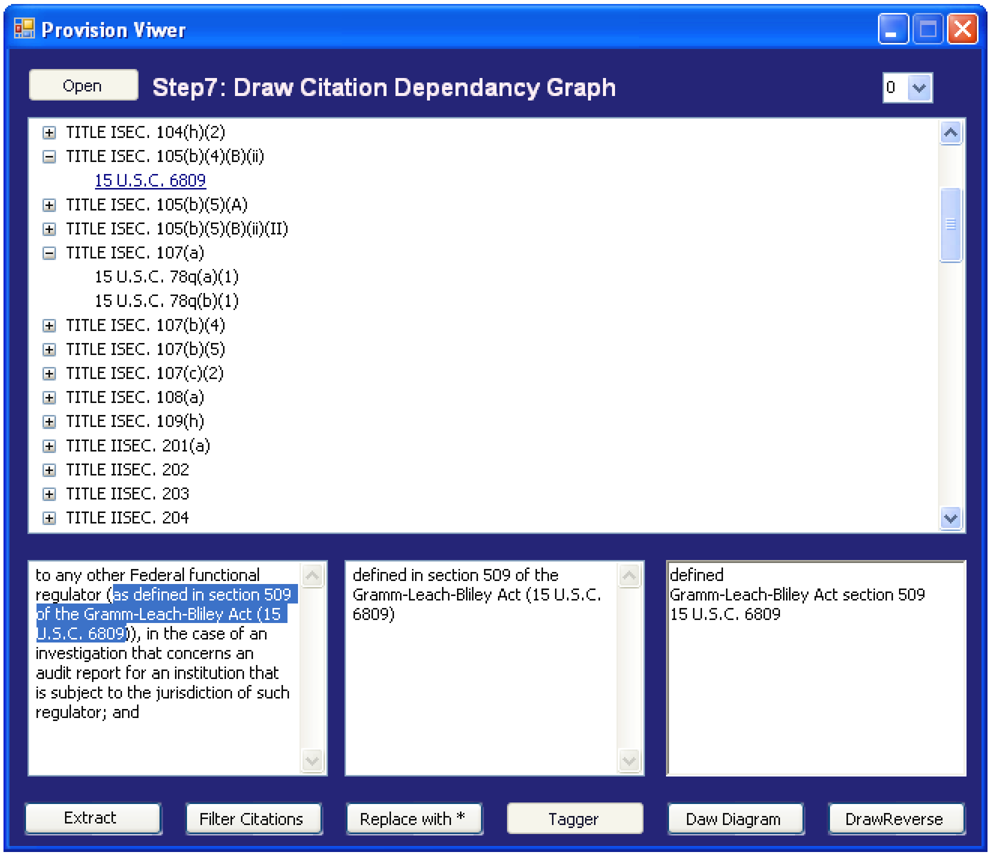


Figure 11. A UML class diagram of the CompDSS data model.

The citation graph builder component (CGBuilder) operates on the object model to construct the citation graph. For this purpose, we implemented the regular expression automata presented in Section 2.3.2 which allows extracting citations that follow the Bluebook and ALWD standards. The user has the flexibility to modify this regular expression to adapt it to other citation style standards. We also implemented the Brill's transformation-based learning tagger algorithm described in [Brill 1992] to be able to perform the part of speech process and extract the verbs that surrounds the citations. Another important aspect of the CGBuilder component is that it accesses the WorldNet library, which free and publicly accessible, to retrieve synonymous words. The verbs and their synonyms are then matched to a list of keywords that we created to be able to classify the citation types into either assertions or amendments. At any time, the user can update the keyword list by adding modifying, or deleting words.

Figure 12 shows a screen snapshot where the provisions extracted with the parser are shown on the main screen. The lower view panels are used for tagging the document to prepare it for the graph building phase. As we can see in Figure 12, the tool also provides an easy way to navigate between provisions, identify the relations between them and give a summary of the provisions. The tree format provided facilitates the analysis and navigation through the provisions of various interrelated laws to understand the content of specific law documents.



Step7: Draw Citation Dependancy Graph

- ⊞ TITLE ISEC. 104(h)(2)
- ⊟ TITLE ISEC. 105(b)(4)(B)(ii)
  - 15 U.S.C. 6809
- ⊞ TITLE ISEC. 105(b)(5)(A)
- ⊞ TITLE ISEC. 105(b)(5)(B)(ii)(II)
- ⊟ TITLE ISEC. 107(a)
  - 15 U.S.C. 78q(a)(1)
  - 15 U.S.C. 78q(b)(1)
- ⊞ TITLE ISEC. 107(b)(4)
- ⊞ TITLE ISEC. 107(b)(5)
- ⊞ TITLE ISEC. 107(c)(2)
- ⊞ TITLE ISEC. 108(a)
- ⊞ TITLE ISEC. 109(h)
- ⊞ TITLE IISEC. 201(a)
- ⊞ TITLE IISEC. 202
- ⊞ TITLE IISEC. 203
- ⊞ TITLE IISEC. 204

to any other Federal functional regulator (as defined in section 509 of the Gramm-Leach-Bliley Act (15 U.S.C. 6809)), in the case of an investigation that concerns an audit report for an institution that is subject to the jurisdiction of such regulator; and

defined in section 509 of the Gramm-Leach-Bliley Act (15 U.S.C. 6809)

defined
Gramm-Leach-Bliley Act section 509
15 U.S.C. 6809

Figure 12. A screen snapshot of the CompDSS tool that shows the extracted provision of the tagging process

The role of the Analyzer component is to compute statistical information about the regulations as well as the citation graphs such as the number of provisions in a regulatory document, the number of provisions that are common between specific laws, the number of provisions on which a particular provision depends on (its outgoing edges), the number of provisions that depend on a particular provision (its incoming edges), etc. The objective is to guide the users when browsing large citation graphs by indicating places of interest based on statistical data.

The Visualization Engine relies on the Graphviz library [Graphviz] to draw the citation graphs. We used various visualization techniques to present the graph in a usable way such as color-coding techniques to distinguish among various regulations, the ability to vary the level of the details displayed in the graph by, for example, changing the tree depth of the provision hierarchical structure, and so on. At a very high-level, the tool can show a graph that exhibits only the name of the studied regulations and the relationship among them. The other extreme will be to show a detailed graph where all the provisions and the relationship among them are displayed. We expect the analysts to vary this level of detail based on the objective of the analysis and their knowledge of the regulations.

2.1. Challenges and Lessons Learned in Developing CompDSS

The development of CompDSS was a challenging task due to the ambiguities and inconsistencies found in law documents. In our work, we spent considerable amount of time understanding law documents, the way they are written, and the standards used in writing such documents. Identifying patterns that can indicated possible conflicts and overlaps was a challenging task to since it required the analysis of multiple regulations that, sometimes, do not pertain to the same area. For example, SOX and HIPAA refer to several common laws although SOX is used to regulate financial reporting and that whereas HIPAA is used to protect the privacy of patient information.

It was also challenging to build CompDSS because it required excellent understanding of the standard citation styles such as the Bluebook and ALWD. Understanding these standards helped build the first prototype of the tool. Applying the prototype on different regulations combined with a trial and error approach was used to harden the first prototype to deal with the ambiguities of regulations. Having a pipe and filter architecture as the generic architectural style for CompDSS made it simple to identify the tool bottleneck which was mainly the visualization engine and the text mining techniques.

Since the focus of this paper was essentially on presenting the citation analysis approach and how it could help detect conflicts and overlaps, we did not deal with scalability, performance, and usability problems; although some of our design decisions, for example, the use of the "neato" rendering techniques instead of "circo" was made with performance in mind. However, we are aware that this poses a serious challenge for the applicability of our approach.

Since the end users of the tool are not necessarily technical people. Another challenge was to make the tool as intuitive as possible, with self-explanatory steps that should facilitate the usability of the tool. However, the large size of graphs makes this task very complex, and there is need to investigate more effective visualization techniques.

4. Case Study

In this section, we show the applicability of our approach by applying it to three U.S. acts, namely SOX, GLBA, and HIPAA, with the objective being to detect conflicts and overlaps among these laws.

4.1. Target Regulations

We briefly present the target regulations used in the case study.

GLBA: The Financial Services Modernization Act, which is also known as the Gramm-Leach-Bliley Act (GLBA) was enacted in November 1999 to allow better synergy among financial institutions including commercial banks, investment companies, brokerage firms, and insurance companies (Kairab, 2004). As a financial act, most of GLBA's seven titles are concerned with financial issues. However, GLBA has another key element which is concerned with privacy and security requirements for personal financial information. Title V of the act focuses on the fact that companies must have a complete security program to protect their customer financial information, as noted in Section 501 of the act "…each financial institution has a continuing obligation to respect the privacy of its customers and protect the security and confidentiality of those customers' nonpublic personal information" (Herrmann, 2007).

GLBA contains several provisions that govern the way organizations should handle the security of customer financial records. For example, GLBA mandates financial institutions to have a comprehensive information security program based on standards to enforce technical and physical safeguards for protecting the information of customers, to ensure confidentiality of customer records, and protect the information from loss, damage, or fraudulent accessibility from none authorized people or parties.

HIPAA: The Health Insurance Portability and Accountability Act (HIPAA) was enacted in August 1996 as part of the healthcare reform during President Clinton's second term to ensure security, privacy and portability of medical insurance and health information (Kairab, 2004). HIPAA consists of five titles that cover the standardization of the healthcare delivery process in an attempt to make it more efficient, setting rules to ensure the privacy of patient information and setting standards for securing this information against abuse or other electronic risks. HIPAA

Title II, in particular, refers to the “Administrative Simplification” provisions that address five security and privacy rules that must be financed by healthcare institutions. These rules ensure proper protection of patient electronic records against attacks and fraudulent use. HIPAA security requirements are mapped to the ISO 17799 standard (Herrmann, 2007).

SOX: The Sarbanes-Oxley Act (SOX), which is also known as the Public Company Accounting Reform and Investor Protection Act was enacted in July 2002. This act was a reaction to a series of corporate scandals due to false reporting of financial and accounting statements (Bhandari & Computing, 2006). SOX requires from organizations to have an internal control framework to monitor the processes that have direct impact of financial reporting. Many aspects of this internal control framework impact the way financial and accounting records are handled from the security perspective. In order for auditors, for example, to be able to assess the adherence to SOX requirements, they must first understand the internal control framework of the organization in question. This requires studying access control of the applications used to ensure authenticity, security and integrity of the information flow between systems, the confidentiality of electronic records, non-repudiation of electronic signatures, etc. The information collected in an internal control report can be seen as requirements for information security programs and can help in security assessment.

4.2. Applying Citation Analysis

The citation graph extracted from SOX, GLBA, and HIPAA is shown in Figure 13. GLBA is represented with a yellow color, SOX in blue, while HIPAA provisions are in red. The objective of the graph is to show how the three acts are interconnected by either referring to each other or to other laws. Due to the size of the graph, it was not possible to show the node labels which represent the provisions identifiers.

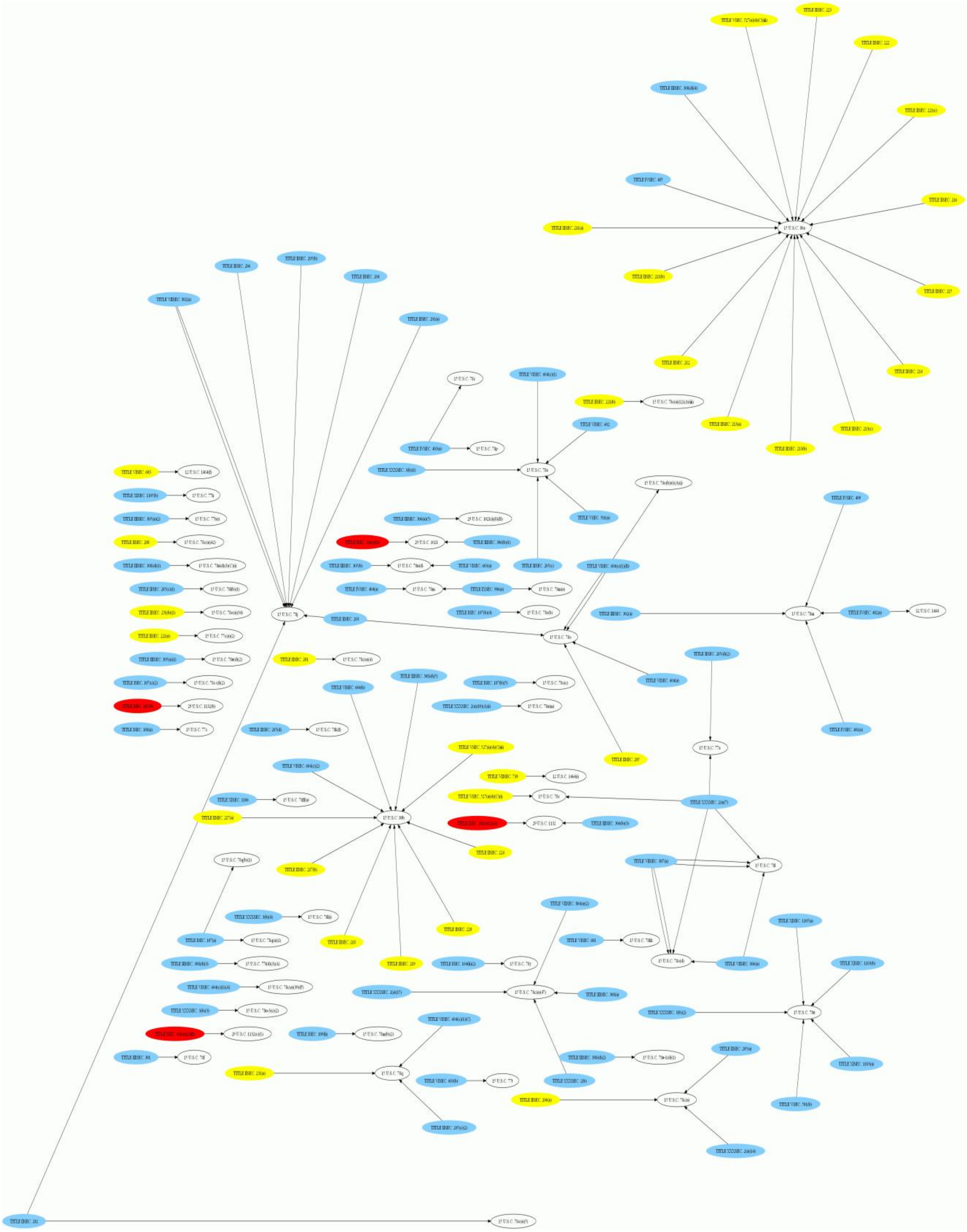


Figure 13. A citation graph that shows the relationship between SOX, GLBA and HIPAA

Table 1 shows a quantitative view depicting the number of laws to which SOX, GLBA, and HIPAA commonly refer. SOX and GLBA refer to four common laws, whereas SOX and HIPAA refer to two laws as shown in the table. Our citation analysis revealed that GLBA and HIPAA do not intersect. We can also see in this table that the Title 15 of the Commerce and Trade Act of the Investor Protection Act Section 78 and Section 80 (15 U.S.C. 78 and 15 U.S.C. 80) are used in 85% of SOX and 91% of GLBA provisions. This is a typical case where, unless the modifications made by SOX and GLBA to the provisions of these laws are carefully monitored, there is high risk of contradictions and duplication of work.

Table 1. Common regulation sections between SOX, GLBA, and HIPAA

Citation	**GLBA**	**SOX**	**HIPAA**
12 U.S.C.1464	2	1	0
15 U.S.C.77	1	7	0
15 U.S.C.78	10	69	0
15 U.S.C. 80	20	5	0
29 U.S.C.1021	0	2	1
29 U.S.C.1132	0	1	3

We analyzed the specific provisions that are referred to by these three regulations and found many cases of duplicated information and potential risks of consistency issues that can lead to conflicts. We chose to report on two cases that are representative of our findings.

Case 1: HIPAA and SOX

The first case is based on the SOX and HIPAA provisions shown in Figure 14. In these provisions we can see two interesting situations where changes made by SOX or HIPAA can affect the other law.

SOX - SEC. 306. INSIDER TRADES DURING PENSION FUND BLACKOUT PERIODS.

(1) IN GENERAL.—Section 101 of the Employee Retirement Income Security Act of 1974 (**29 U.S.C. 1021**) is amended by ***redesignating*** the second subsection (h) as subsection (j), and by ***inserting*** after the first subsection (h) the following new subsection:

''(i) NOTICE OF BLACKOUT PERIODS TO PARTICIPANT OR BENE
FICIARY UNDER INDIVIDUAL ACCOUNT PLAN.—

. . .

(3) CIVIL PENALTIES FOR FAILURE TO PROVIDE NOTICE.— Section 502 of such Act **(29 U.S.C. 1132)** is amended—

(A) **in subsection (a)(6),** by ***striking*** ‘‘(5), or (6)’’ and ***inserting* ‘‘(5), (6), or (7)**’’;

(B) by redesignating paragraph (7) of subsection (c) as paragraph (8); and

(C) by inserting after paragraph (6) of subsection (c) the following new paragraph:

‘‘(7) **The Secretary may assess a civil penalty against a plan administrator of up to $100 a day from the date of the plan administrator’s failure or refusal to provide notice to participants and beneficiaries in accordance with section 101(i).** For purposes of this paragraph, each violation with respect to any single participant or beneficiary shall be treated as a separate violation.’’.

HIPAA- SEC.101. REPORTING AND ENFORCEMENT WITH RESPECT TO CERTAIN ARRANGEMENTS.—

. . .

(b) ENFORCEMENT WITH RESPECT TO HEALTH INSURANCE ISSUERS.—

Section 502(b) of such Act (29 U.S.C. 1132(b)) is amended by ***adding*** at the end the following new paragraph:

‘‘**(3) The Secretary is not authorized to enforce under this part any requirement of part 7 against a health insurance issuer offering health insurance coverage in connection with a group health plan (as defined in section 706(a)(1)). Nothing in this paragraph shall affect the authority of the Secretary to issue regulations to carry out such part.**’’.

. . .

(e)

(1) IN GENERAL.—Section 101 of such Act (**29 U.S.C. 1021**) is amended—

(A) by ***redesignating*** subsection (g) as subsection (h), and

(B) by ***inserting*** after subsection (f) the following new subsection:

‘‘(g) REPORTING BY CERTAIN ARRANGEMENTS.—The Secretary may, by regulation, require multiple employer welfare arrangements providing benefits consisting of medical care (within the meaning of section 706(a)(2)) which are not group health plans to report, not more frequently than annually, in such form and such manner as the Secretary may require for the purpose of determining the extent to which the requirements of part 7 are being carried out in connection with such benefits.’’.

(2) ENFORCEMENT.—

(A) IN GENERAL.—Section 502 of such Act **(29 U.S.C.1132)** is amended—

(i) **in subsection (a)(6),** by ***striking*** ‘‘under subsection (c)(2) or (i) or (l)’’ and ***inserting*** ‘‘under paragraph **(2), (4), or (5)** of subsection (c) or under subsection (i) or (l)’’; and

(ii) in the last 2 sentences of subsection (c), by striking ‘‘For purposes of this paragraph’’ and all that follows through ‘‘The Secretary and’’ and inserting the following:

‘‘(5) **The Secretary may assess a civil penalty against any person of up to $1,000 a day from the date of the person’s failure or refusal to file the information required to be filed by such person with the Secretary under regulations prescribed pursuant to section 101(g).**

‘‘(6) The Secretary and’’.

Figure 14. Provisions taken from SOX and HIPAA to illustrate the fist case of overlap and possible conflicts

The first situation is with respect to the relationship between provision 306(b)(3) of SOX and provision 101(e)(2)(A) of HIPAA that is depicted in the citation graph shown in Figure 15 and which corresponds to the Amend-Amend pattern. It consists of the fact both these laws amend Section 502 of The Employee Retirement Income Security Act of 1974 (code 29 U.S.C. 1132). The amendment consists in both cases of striking and inserting paragraphs, in particular to Subsections (c) and (a)(6) of Section 502.

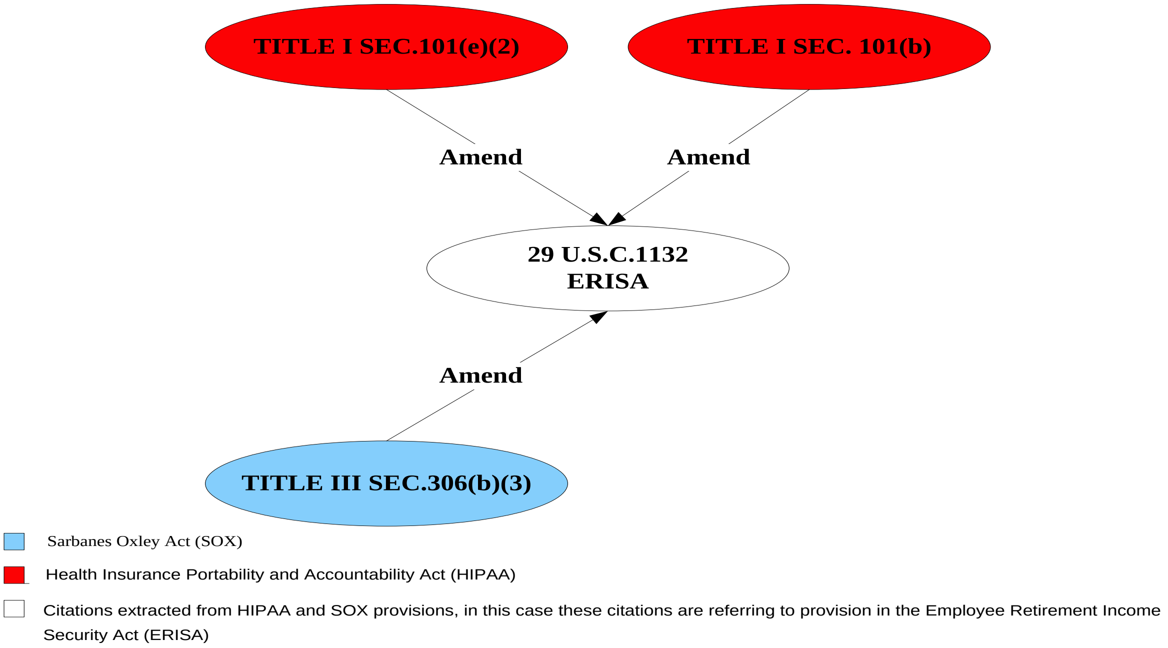

TITLE I SEC.101(e)(2)
TITLE I SEC. 101(b)
Amend
Amend
29 U.S.C.1132
ERISA
Amend
TITLE III SEC.306(b)(3)
Sarbanes Oxley Act (SOX)
Health Insurance Portability and Accountability Act (HIPAA)
Citations extracted from HIPAA and SOX provisions, in this case these citations are referring to provision in the Employee Retirement Income Security Act (ERISA)

Figure 15. A graph showing SOX and HIPAA provisions amending the same provision of the 29 U.S.C 1132 law

Although SOX and HIPAA agree on the rights and obligations of the Secretary, who is defined as an officer of the health and human services, in SOX, the secretary has a wider authority than in HIPAA. This is indicated by the fact that SOX added a new paragraph (7) to the Employee Retirement Income Security Act of 1974 by allowing the Secretary to assess a civil penalty against a plan administrator who refuses or fails to provide notice to participants and beneficiaries, which is not the case in HIPAA.

The second situation, which is illustrated in the citation graph of Figure 16, shows an example of an overlap between SOX and HIPAA with respect to Section 101 of ERISA (29 U.S.C. 1021). Both HIPAA and SOX amend the same section but two different subsections. Although this decreases the possibility of a conflict, it enforces the fact that there is a strong relation between the policies that HIPAA and SOX mandate in these provisions. In this case, both regulations are adding new obligations on the plan administrator regarding the disclosure and reporting of information. SOX Section 306(b) amended Section 101(i) of ERISA (29 U.S.C. 1021) to ensure that plan administrators notify participants at least 30 days in advance of “blackout periods” (a temporary period of at least 3 days during which contracts or policies are suspended). On the other hand HIPAA Section 101(e)(1) amended Section 101of ERISA (29 U.S.C. 1021) by adding a new paragraph to establish a filing requirement that was required by the Multiple Employer Welfare Arrangements (MEWA). This requirement consists of a new obligation on plan administrators that mandates reporting and filing annually to The Employee Benefits Security Administration (EBSA) information related to the requirements in Part 7 of ERISA which is added by HIPAA and other similar regulations such as Mental Health Parity Act to ensure compliance by MEWA. These requirements are not needed for SOX compliance.

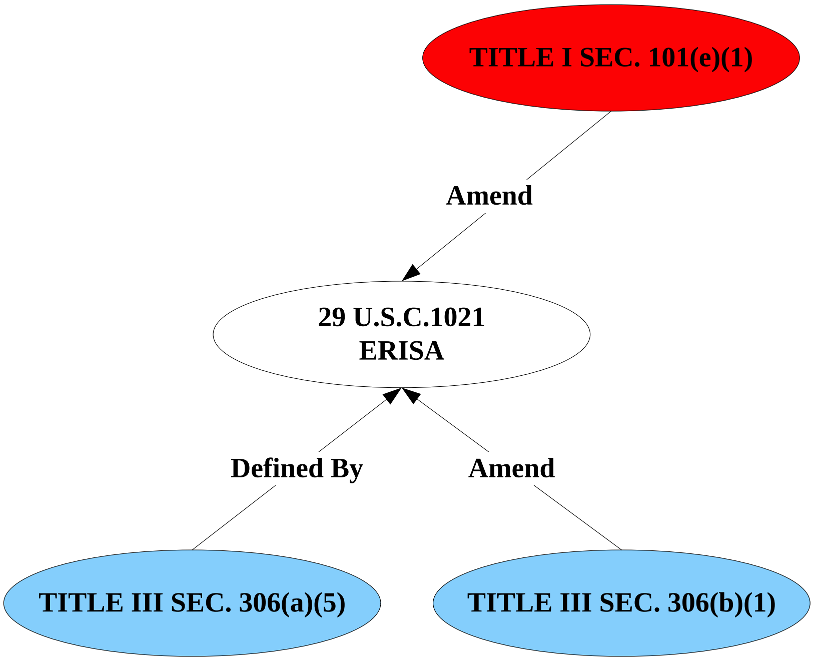


Figure 16. An example of SOX and HIPAA provisions amending the same provision of the 29 U.S.C 1132 law but different sections

Case 2: SOX and GLBA

SOX and GLBA overlap by addressing record keeping and the protection of investors' private information. The relationship between SOX provision 205(c)(2) and GLBA 231(a) (shown in Figure 17 and the corresponding graph in Figure 18) consists of the fact that they both amend the same provision of Section 17 of the Securities Exchange Act of 1934 (code 15 U.S.C 78q). This is again falls into the Amend-Amend pattern. Before the modification made by SOX to this section, it was sufficient from GLBA’s perspective for an accountant that certifies the financial documents and reports of an organization to be an independent public account. SOX, on the other hand, amended Section 17 of the Securities Exchange Act of 1934 in such a way that the accountant must be working in a firm which must be registered by the Public Company

Accounting Oversight Board (PCAOB) (PCAOB, 2009). PCAOB is a private non-profitable corporation created by SOX to oversee the auditors of public companies to protect the interest of investors (PCAOB, 2009). This amendment made by SOX and which affected GLBA requires that all tool features, audit processes, and organizational policies that were compliant with the previous version of GLBA be updated.

SOX - TITLE II—AUDITOR INDEPENDENCE

SEC. 205. CONFORMING AMENDMENTS

(c) OTHER REFERENCES.—The Securities Exchange Act of 1934 (15 U.S.C. 78a et seq.) is amended—

(1) in section 12(b)(1) (15 U.S.C. 78l(b)(1)), by striking ''independent public accountants'' each place that term appears and inserting ''a registered public accounting firm''; and

(2) in subsections (e) and (i) of section 17 (15 U.S.C. 78q), by striking ''an independent public accountant'' each place that term appears and inserting ''a registered public accounting firm''.

GLBA -Subtitle C—Securities and Exchange Commission Supervision of Investment Bank Holding Companies

SEC. 231. SUPERVISION OF INVESTMENT BANK HOLDING COMPANIES

BY THE SECURITIES AND EXCHANGE COMMISSION.

(a) AMENDMENT.—Section 17 of the Securities Exchange Act of 1934 (**15 U.S.C. 78q**) is amended—

(1) by redesignating subsection (i) as subsection (k); and

(2) by inserting after subsection (h) the following new subsections:

''(i) INVESTMENT BANK HOLDING COMPANIES.—

''(1) **Elective supervision of an investment bank holding company not having a bank or savings association affiliate**

…

" (3) Supervision of investment bank holding companies

"(A) Record keeping and reporting

…

"(ii) Form and contents Such records and reports shall be prepared in such form and according to such specifications (**including certification by a registered public accounting firm**), as the Commission may require and shall be provided promptly at any time upon request by the Commission. Such records and reports may include—

Figure 17. Provisions taken from SOX and GLBA to illustrate a case of overlap that can lead to a conflict unless careful monitoring of the changes is performed

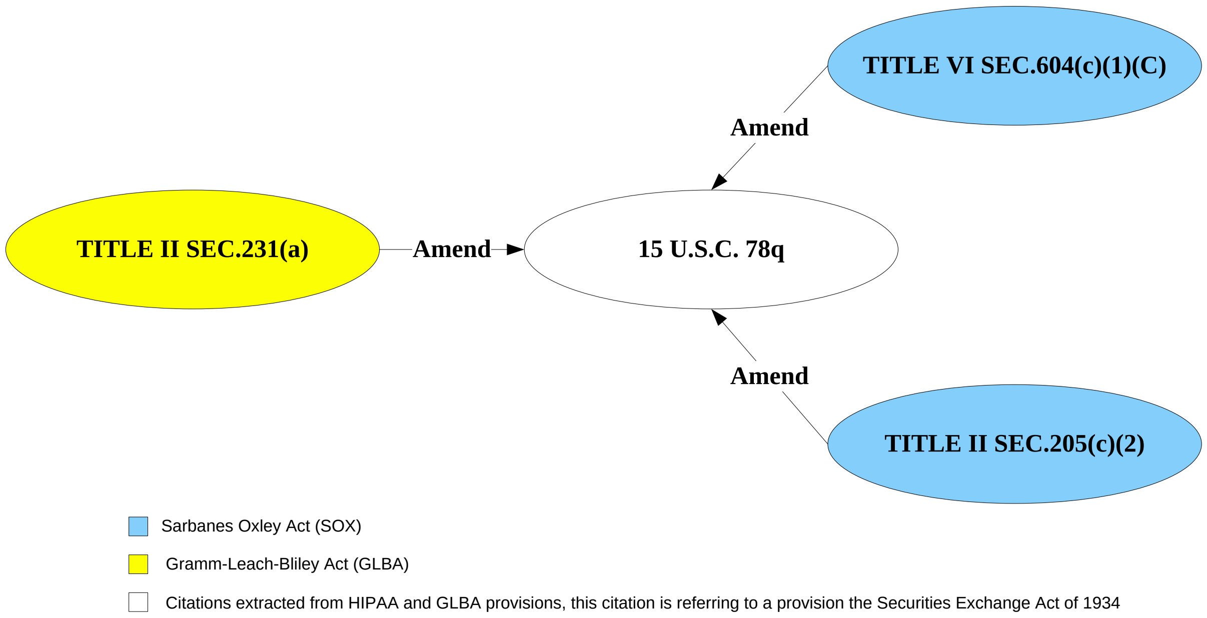

TITLE VI SEC.604(c)(1)(C)
Amend
TITLE II SEC.231(a)
Amend
15 U.S.C. 78q
Amend
TITLE II SEC.205(c)(2)
Sarbanes Oxley Act (SOX)
Gramm-Leach-Bliley Act (GLBA)
Citations extracted from HIPAA and GLBA provisions, this citation is referring to a provision the Securities Exchange Act of 1934

Figure 18. Example of SOX and GLBA provisions amending the same provision “15 U.S.C 78q”

5. Related Work

There exist several studies that use citation analysis to extract relevant information from different documents such websites, or scientific publications [He 2002, Zhao 2002]. The proposed techniques, however, operate on documents that are completely different from law documents in term of their structures, and the citation standards that are used, which make them impractical in the context of our research.

Breaux proposed the Frame-Based Requirements Analysis Method (FBRAM) to obtain legal requirements from U.S. regulatory documents [Breaux 2009]. His method is intended to help software engineers understand legal requirements in order to avoid non compliance when building software systems where compliance requirements can turn into system functional requirements. The author studied the sources of ambiguities in legal document, such as the effect of facts, definitions, constraints, the language ambiguity and cross-references. His study of the effect of cross-references, however, was limited to defining a set of patterns to extract internal references, to support mappings between legal requirements [Breaux 2009]. In our study, we use citation analysis to help analysts work with multiple regulations by understanding and keeping track of the changes made by laws and that may affect other laws.

The objective of the REGNET project is to build a formal information infrastructure for regulatory information management and compliance assistance to supports U.S. federal and state regulations [Lau 2006a, Lau 2006b]. One of the main outcomes of the project is a set of techniques for locating and comparing related regulations based on information retrieval techniques, feature matching, etc. (Lau, Law, & Wiederhold, 2006) (Lau, Wang, & Law, 2006). Although the work of the REGNET approach does not focus on citation analysis, we believe that their techniques complement the one presented in this paper.

Jacobson et al. developed an Interactive Legal Citation Checker, a proprietary tool to extracts citations from legal documents [Jacobson 2006]. The aim was to help authors of legal documents write citations that comply with a citation style. This tool, however, only extracts the citations and does not detect the various relations among citations such as the ones presented in this paper (i.e., assertion and amendments).

The Semantics-Based Legal Citation Network is a prototype visualization tool that focuses on case law citations. The tool was proposed by Zhang et al. [Zhang 2007]. The goal is to help attorneys and other people who work on the legal field study cases and legal issues without the need to go through the whole cases or the daunting task of manual citation search. The citation viewer focuses on one legal issue and allows an attorney to navigate (forwards and backwards) through the citation network to study how the issue was developed and handled in different cases over the years. However, their tool helps simply navigate through regulatory documents and does not focus on detecting overlaps or conflicts among these documents.

6. Conclusion and Future work

In this paper, we argued that there is a need to investigate techniques and tools to help software companies deal with a large number of regulatory compliance requirements. This is because many software systems they develop are for organizations that are heavily regulated and must follow various local and international laws, which dedicate key features that these software systems must support or even the process by which they are to be built.

We discussed how citation analysis can be used to understand and analyze multiple regulations as well as detect overlaps and possible conflicts among regulations. We also presented our citation analysis approach, which involves building a citation graph by parsing the content of multiple regulations to identify the provisions, citations, and the types of relations among citations. We discussed the fact that amendments (one of the relations among the citing and the cited provisions) represent risks of conflict. We suggested that this type of relation should be carefully studied.

We also presented a prototype tool called CompDSS that supports citation analysis. Although many features of the tool are still under development, we were able to use it to detect overlaps and risks of conflicts among three laws, namely, SOX, HIPAA, and GLBA, the regulations used in our case study.

Although the approach presented in this paper and the results of the case study are promising, there are many issues that need to be addressed:

- Citation analysis cannot be applied to analyze the dependencies among laws from different countries since it is unlikely that these laws refer to the same common laws. This limits our approach from being used by corporations that wish to apply it to detect overlaps and conflicts at the international level. We do not have the solution to this problem for the time being.
- Our process for extracting citations relies on citation style standards. However, there might be many citations that do not rely on the citation style standards. One possible solution to this issue is to use text mining techniques to extract such citations. We intend to further investigate this path.
- The size of a citation graph that involves many laws can be relatively large. This can hinder the effective analysis of these laws and the relationships among them. What we need is to investigate ways to improve the usability of these graphs by adding support of various visualization techniques that have been shown to be effective in other areas of software engineering.
- CompDSS follows a pipe and filter architectural style, each phase depends on the previous one. The bottleneck of the tool is the visualization engine which is based on the Graphviz package and libraries. Citation graphs could become extremely large, and may hinder the applicability of our approach, unless we work towards defining better algorithms for improved performance and scalability.

- In this paper, the detection of possible overlaps and conflicts is a user-intensive task. For example, it is up to the analyst to find places where various amendments of the same provisions occur. In addition to this, we recognize that our pattern library does not cover all possible cases of potential conflicts and overlaps. Analyst working in particular areas of law may need to define new patterns that are most likely to occur. This will require careful analysis of law documents.
- We need to work with users working with multiple regulations to be able to improve our approach as well as the tool.
- Finally, we also need to compare our approach with existing studies such as the one presented by the members of the REGNET project, in which information retrieval techniques have been used to detect relationships among law documents.

7. References

Antoniou 1999 Antoniou, G., Billington, D., Maher, M., "On the Analysis of Regulations using Defeasible Rules", *In Proc. of the 32nd Hawaii International Conference on Systems Science,* pp. 225-225, 1999.

AALL 2002 *The AALL Universal Citation Guide 2.1.* by The American Association of Law Libraries Citation Format Committee, 2002.

Bhandari 2006 Bhandari, S., Computing, T., "Information Trust and Compliance Issues", *In Proc. of* the *MS Technology Management - Information Security and Management* , pp. 1-15, 2006.

Bagley 2006 Bagley, C. E., *Strategic Compliance Management.* Harvard Business Publishing, 2006.

Barris 2007 Barris, L. J., *Understanding and Mastering the Bluebook: A Guide for Students and Practitioners,* Carolina Academic Press, 2007.

Breaux 2009 Breaux, T., "Legal Requirements Acquisition for the Specification of Legally Compliant Information Systems", *Ph.D. Dissertation, North Carolina, USA: North Carolina State University,* 2009.

Brill 1992 Brill, E., "A simple rule-based part of speech tagger", *In Proc of the 3rd Conference on Applied natural language processing,* pp. 152-155, 1992.

Brill 1995 Brill, E., “Transformation-based error-driven learning and natural language processing: a case study in part-of-speech tagging”, *Computational Linguistics, 21* (4), pp. 543-565, 1995.

CGULC 2006 The *Canadian Guide to Uniform Legal Citation,* McGill Law Journal, Toronto: Thomson Carswell, 2006, http://lawjournal.mcgill.ca/citeguide.php

Chang 2005 Chang, H., Kim, K., “Design of Inside Information Leakage Prevention System in Ubiquitous Computing Environment”, *Lecture Notes in Computer Science, vol. 3483,* pp. 128-137, 2005.

Dickerson 2003 Dickerson, D, Association of Legal Writing Directors, *ALWD citation manual : a professional system of citation,* New York: Aspen, 2003.

Ehrlich 1922 Ehrlich, E., “The Sociology of Law”, *Harvard Law Review, 36 (*2), pp. 130-145, 1922.

Ernst 2006 The Ernst & Young 2006 Global Information Security Survey: URL: http://www.ey.com/Publication/vwLUAssets/GISS_India_release_2006/$file/GISS_India_release_2006.pdf

Gamma 1994 Gamma E., Helm R., Johnson R., and Vlissides J., *Design Patterns Elements of Reusable Object-Oriented Software*, Addison Wesley Professional, 1994.

Graphviz Graphviz: Graph Visualization Software. Retrieved Augest 11, 2009, from http://www.graphviz.org/

Hamdaqa 2009 Hamdaqa, M., Hamou-Lhadj, A. “Citation Analysis: An Approach for Facilitating the Understanding and the Analysis of Regulatory Compliance Documents”, *In Proc. of the 6th International Conference on Information Technology: New Generations*, pp. 278-283, 2009.

Hamou-Lhadj 2007 Hamou-Lhadj, A-K, Hamou-Lhadj, A., “Towards a Compliance Support Framework for Global Software Companies”, *In Proc. of the IASTED Conference on Software Engineering, 2007.*

Herrmann 2007 Herrmann, D. S., *Complete Guide to Security and Privacy Metrics: Measuring Regulatory Compliance, Operational Resilience, and ROI,* Auerbach Publications, 2007.

Jacobson 2006 Jacobson, R. L., “Interactive legal citation checker”, U.S. Patent No. 7028259, 2006.

Kairab 2004 Kairab, S., *A practical guide to security assessments,* Auerbach Publications, 2004.

Lau 2006a Lau, G., Law, K., & Wiederhold, G., "A relatedness analysis of government regulations using domain knowledge and structural organization", *International Journal of Information Retrieval , 9* (6), pp. 657-680, 2006.

Lau 2006b Lau, G., Wang, H., & Law, K., "Locating related regulations using a comparative analysis approach", *In Proc. of the International Conference on Digital Government Research*, pp. 229-238, 2006.

Miller 1995 Miller, G. A., "WordNet: a lexical database for English", *Commununications of the ACM , 38 (11),* pp. 39-41, 1995.

Ouellet 2009 Ouellet, E., Proctor, P. E., "*Magic Quadrant for Content Monitoring and Filtering and Data Loss Prevention",* Gartner RAS Core Research Note G00168012, 2009.

PCAOB The Public Company Accounting Oversight Board Mission: URL: www.pcaobus.org

Silverman 2008 Silverman, M. G., *Compliance Management for Public, Private, Or Nonprofit Organizations,* McGraw-Hill Professional, 2008.

Suber 1999 Suber, P., "Amendment", An essay that appeared in *Philosophy of Law: An Encyclopedia*, pp. I.31-32. Garland Publications, 1999.

Tarantino 2006 Tarantino, A., *Manager's Guide to Compliance Sarbanes-Oxley, COSO, ERM, COBIT, IFRS, BASEL II, OMB's A-123, ASX 10, OECD Principles, Turnbull Guidance, Best Practices, and Case Studies*, Wiley, 2006.

Zhang 2007 Zhang, P., Koppaka, L., "Semantics-based legal citation network", *In Proc. of the 11th International Conference on Artificial Intelligence and Law*, pp. 123-130, 2007.

He 2001 He, Y. and Hui, S. C., "Mining a web citation database for author co-citation analysis", *Inf. Process. Manage. 38(2),* pp. 491-508, 2002.

Zhao 2002 Zhao, D. and Logan, E., "Citation analysis using scientific publications on the Web as data source: A case study in the XML research area" *Scientometrics, 54(3),* pp. 449-472, 2002.

Oppliger 2000 Rolf Oppliger, "Privacy protection and anonymity services for the World Wide Web (WWW)", *Journal of Future Generation Computer Systems, 16(4), Elsevier*, pp. 379-391, 2000.

Chang 2004 Beom-Hwan Chang, Dong-Soo Kim, Hyun-Ku Kim, Jung-Chan Na, Tai-Myoung Chung, "Active security management based on Secure Zone Cooperation", *Future Generation Computer Systems, 20(2), Elsevier,* pp. 283-293, 2004.