McLaren, B.M. & Ashley, K.D. (2011). Can temporal representation and reasoning make a difference in automated legal reasoning? Lessons from an AI-based ethical reasoner. In the *Proceedings of the 13th International Conference on Artificial Intelligence and Law (ICAIL-2011)*.

Can Temporal Representation and Reasoning make a Difference in Automated Legal Reasoning? Lessons from an AI-based Ethical Reasoner

Bruce M. McLaren
Carnegie Mellon University
Human-Computer Interaction Institute
Pittsburgh, PA, USA
001-412-268-8278
bmclaren@cs.cmu.edu

Kevin D. Ashley
University of Pittsburgh
Intelligent Systems Program
Pittsburgh, PA, USA
001-412-624-7496
ashley@pitt.edu

ABSTRACT
Given a renewed interest in the field of AI and Law in more complex factual representations of legal cases in terms of narratives, techniques for representing and reasoning about temporal orderings of facts will become increasingly important. The SIROCCO (**S**ystem for **I**ntelligent **R**etrieval of **O**perationalized **C**ases and **CO**des) program employed a representation for the temporal ordering of events in ethics cases in a way that informed determinations of whether and how ethical norms were violated and if the problem and other cases were normatively analogous at a deeper level. At the same time, the program supported ordinary case enterers in translating the facts of textually described cases into a machine-processable representation. This paper presents these previously unpublished aspects of the work including a report of an empirical evaluation of the contribution of the temporal representation to the program's success in retrieving relevant norms and cases. Although the results were negative, a consideration of the reasons why is illuminating. While SIROCCO dealt with engineering ethics cases, it is clear that similar temporal considerations apply in legal cases and that the approach is likely to be useful in legal narrative representations.

Keywords
Narrative case models; representing temporal ordering; case similarity.

1. INTRODUCTION
Two recent developments in the field of AI and Law focus attention on the importance of including a temporal component in representing legal cases. The first is simply a renewed interest in developing more complex representations of facts that go beyond treating them as lists of factors (see, e.g., [29]; and the recent series of three Workshops on Modelling Legal Cases and Legal Rules 2010). The second is the emergence of renewed interest in narrative models of the events in legal cases [28; 5] and with it a realization of the importance of the temporal sequence of events for assessing similarity among cases. As Walton has put it:

> There is something about the common sequence of events that makes the one case similar to the other. First the plaintiff bought some product that he … thought … was reasonably safe to use. Then … when he used the product this defect caused some harm that impacted badly on his health. There is a thread, or sequence of events that is of the same kind in both cases. It started in the same way, went through the same kind of chain of events, and ended in the same way. [28, p. 228]

If automated similarity assessments are to penetrate surface features and identify the deeper shared features that have normative significance, representing the temporal sequence of events is key. Again, quoting Walton:

> If we distinguish, following Ashley (2009), between deep and shallow analogies, a template that matches up the same sequence fitting two cases can reveal a deep similarity that is more significant, as opposed to a shallow similarity in which the two cases do not appear to be similar. [28, p. 229]

Given the renewed interest in narrative models and temporal event sequences, it follows there should be an increasing interest in how best to represent time and the temporal ordering of events in cases for purposes of similarity assessment. Of course, AI and Law researchers have taken into account the normative implications of time in a number of contexts, for example: (1) important statutory dates (of adoption, amendment, effective dates, etc. (See, e.g., [23]). (2) dates of precedents (and their effects on *stare decisis* and on trends in meaning of legal concepts. See, e.g., [4; 25]. (3) dates of events in a case's facts with respect to what laws were applicable at the time of the events in question (See, e.g., [7]).

In this paper, however, we are concerned with the timing of events in a case's facts in a different sense: assuming one knows or can determine the laws applicable at the time, how does the time ordering of the factual events affect whether those laws were violated and whether other apparently similar cases are indeed relevant in Walton's sense?

In his SIROCCO program, McLaren developed a representation of the temporal ordering of events in ethics cases in a way that informed determinations of whether and how ethical norms were violated and if the problem was normatively analogous in a deeper sense to other cases. In addition, he empirically evaluated the contribution of the temporal representation to the program's success in retrieving relevant norms and cases. Although reported

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ICAIL '11, June 6-10, 2011, Pittsburgh, PA
Copyright © 2011 ACM 978-1-4503-0755-0/11/06 ... $10.00

in the dissertation, the results of that experiment have never been published.

2. SIROCCO, THE COMPUTATIONAL MODEL

McLaren used AI techniques, in particular case-based reasoning and A* heuristic search, to investigate how experienced ethical decision-makers apply ethics principles and past cases and the effects of that application in terms of fleshing out the meanings of the norms [17; 19]. The work focused on the code of ethics of the National Society of Professional Engineers (NSPE) and hundreds of cases decided by the NSPE's Board of Ethical Review (BER) [22]. Each case included a description of the facts and explanations of how and why the code provisions apply (or not) and how the cases are similar to or different from past cases. He found evidence that, in the process of explaining their decisions, these experienced ethical decision-makers tended to define engineering ethics principles and past cases *extensionally* (i.e., they operationalized them). For instance, the Board instantiated code principles by linking them to clusters of questioned and critical facts in a case, grouped related code principles together in the context of specific cases, and instantiated past cases by linking them to clusters of questioned and critical facts and by analogizing or distinguishing them.

The SIROCCO program applied information associated with these operationalization techniques to improve retrieval of relevant ethics code provisions and past cases from the NSPE's on-line database of cases. The ethics problem facts were represented as chronologically ordered sets of steps expressed in a limited language called ETL (Ethics Transcription Language) that provided a vocabulary of fact primitives and time qualifiers.

The main results in this dissertation have been published, most comprehensively in [16]. An experiment demonstrated that SIROCCO's computational model of a critical subset of the operationalization techniques made more accurate predictions of the facts, ethics code principles and past cases that are relevant to analyzing new problems than various alternative methods, including two full-text retrieval approaches. An ablation study compared SIROCCO with a version in which the core set of operationalization techniques was turned off, demonstrating the specific contribution of these techniques to the program's retrieval effectiveness and providing strong evidence of their effect in supplying an extensional meaning for the general ethics code principles and past cases. The effect was statistically significant; in two related experiments (involving exact and more inexact matching criteria), operationalization information accounted for 33% and 52% of SIROCCO's combined precision and recall (i.e., F-measure) results [15; 16; 18; 3].

The foci of this paper, however, are on the representation and use of temporal knowledge in SIROCCO, an experiment to assess the contribution of this temporal knowledge to case retrieval efficacy, and the previously unpublished results of *that* experiment.

3. NORMATIVE/TEMPORAL EXAMPLES

While it is not surprising that temporal information plays a role in normative reasoning, it is instructive to examine some examples in detail. Our survey of the NSPE cases confirmed that temporal information plays a sometimes crucial role in the decision-makers' interpretation of the open-textured concepts of the ethics code provisions, in mapping the facts of cases to the code concepts, and in defining relevant similarities and differences among cases.

For example, the Board used temporal reasoning in deciding whether it was ethical for an engineer, B, to distribute a brochure listing another engineer, A, as a key employee after B had notified A of A's termination but before A was actually terminated (Case 83-1-2) and to distribute the brochure after A was actually terminated (Case 83-1-3). In applying Code II.5.a., which states that "brochures or other presentations incident to the solicitation of employment shall not misrepresent pertinent facts concerning employers, employees, associates . . . with the intent and purpose of enhancing their qualifications and their work," the Board found a violation in 83-1-2 where "Engineer B distributed the brochure while Engineer A was still employed but had been given a notice of termination by Engineer B. That could easily mislead potential clients into believing that Engineer A, noted as a key employee, would be available in the firm for consultation on future projects. Moreover, Engineer B distributed the brochure after Engineer A had left the firm." In applying Code III.3.a., which states that "Engineers shall avoid use of statements containing a material misrepresentation of fact or omitting a material fact necessary to keep statements from being misleading; statements intended or likely to create an unjustified expectation....", the Board said that "during the interim period between Engineer A's being given notice of termination and his actual cessation of employment, Engineer B had an obligation, during negotiations with a prospective client, to inform the client of Engineer A's pending termination. However, once Engineer A had been formally dismissed, Engineer B had an ethical obligation to cease using the brochure with Engineer A's name in it entirely."

Other examples where temporal knowledge is used to determine how ethics codes apply include: (1) deciding in Case 89-5-1 that it was not true that "a sufficient amount of time has passed … to dilute the ethical obligations owed by Engineer A to his former client, the contractor" resulting in continuing "conflicts of interest to their … clients" under Code II.4.a. and related codes. (2) deciding in Case 76-5-1 that "it is clear that when Engineer B entered the scene, Engineer A had no contract with PDQ and negotiations had been terminated [and] thus the client was free to turn to another engineer despite a Code 11(a) prohibiting engineers from attempting to supplant "another engineer in a particular employment after becoming aware that definite steps have been taken toward the other's employment."

The Board also uses temporal reasoning to discern relevant similarities and differences among cases. In Case 83-1-1 (factually related to 83-1-2 and 83-1-3 above) in deciding that Engineer A acted unethically in soliciting B's clients while still in B's employ, the Board distinguished, based on mismatches in the temporal relationships, a superficially similar past case (77-11-1), where the engineers were found not to be unethical in soliciting their former employer's former clients. The cases were normatively different because: (1) 83-1-1 involved the solicitation of *current* clients versus *former* clients in Case 77-11-1, and (2) the engineer in Case 83-1-1 *was still* employed when he solicited these clients, while the engineers in Case 77-11-1 *were not still* employed.

In two other superficially similar conflicts of interest cases, the Board cites Code III.1.c as applying to Case 62-7-1 but does not do so in their analysis of Case 91-6-1, apparently based on temporal differences. The Code states, "Engineers shall not accept outside employment to the detriment of their regular work or interest. Before accepting any outside employment they will notify their employers." The intent of this code is to avoid a special type of conflict of interest that arises when an engineer has an on-going job or assignment and at the same time accepts other work that may be detrimental to the regular assignment. In Case 91-6-1, Engineer A's employment with the corporation began long after his contract with the agency concluded, and the Board concluded that engineer A's work for the agency was not "regular work or interest". In Case 62-7-1, Engineer Z has a "regular" assignment for a governmental body and, at the same time, accepts a conflicting assignment with a private company.

Although SIROCCO's domain of expertise is engineering ethics, these examples confirm Walton's observation that analogous issues of temporal dependence arise in legal contexts. Legal obligations to warn of risks or to disclose information usually do not arise until the purported obligee learns of the danger or the information herself. A party cannot accept an offer until the offer has been made. Wherever narrative representations of legal cases are processed, temporal orderings of events will be important.

4. REPRESENTING / USING TEMPORAL KNOWLEDGE IN SIROCCO

SIROCCO accepts new cases in the aforementioned case-representation language, the *Ethics Transcription Language* (ETL). The language represents the actions and events of a scenario as an ordered list (i.e., a *Fact Chronology*) of individual sentences (i.e., *Facts*), each consisting of (1) *Actors and objects*, instances of general actors and objects which appear in the scenario, (2) a *Fact Primitive,* the action or event in which the actor and/or object instances participated, and (3) a *Time Qualifier,* a temporal relation that specifies how a Fact relates to other Facts in time. A predefined set of general Actor, Object, Fact Primitive, and Time Qualifier types are used in the representation. At least one Fact in the Fact Chronology is designated as the *Questioned Fact*; this is the action or event corresponding to the ethical question raised in the scenario.

Cases stored in SIROCCO's case base are represented in the *Extended Ethics Transcription Language* (EETL), of which ETL is a subset. EETL provides all of the elements described above but, in addition, provides a template and standard components for representing the Board's analysis of a case, including its conclusion (i.e., ethical, unethical, or undecided), the protagonist whose action is questioned, the Board's general argument structure, a linking of the critical facts of the case to the citations supported by those facts, and a series of other details, such as which citations are more important and which citations are grouped together. In essence, EETL models arguments as a series of code and case operationalizations, some of which support the Board's conclusion, some of which conflict with the conclusion, and some of which simply provide relevant background information. Two such operationalizations, *Code Instantiations* and *Case Instantiations,* are defined by the case enterers based on their interpretation of the Board's rationale in citing a code provision or a case. An Instantiation relates a questioned fact, certain critical facts, and the temporal sequence of those facts to the citation of a code provision or past case. Using the instantiations to guide matching makes the retrieval algorithm more efficient.

The case base of SIROCCO is comprised of 184 *foundational cases*, covering 135 fact situations[1] and culled from the grand total of 475 cases decided by the NSPE BER between 1958 and 1992. These cases were used to design, implement, and refine the program. All 75 of the NSPE BER ethics code provisions are also encoded in SIROCCO. The goal was to include all of the cases covering a reasonable number of important ethics topics or codes (e.g., public safety) but also to provide some (minimal) coverage of cases outside of those topics. As such, the foundational case base includes all 135 of the cases that cite at least one code related to at least one of the following selected topics: public safety, confidential information, duty to employer, credit for engineering work, proprietary interests, and honesty in reports and public statements. The foundational case base also includes 49 cases that do not cite any codes from the selected topics. The foundational cases were represented with the support of a case-acquisition web site and were transcribed into EETL by a total of 12 independent case enterers (to provide objectivity for our experiments).

To retrieve relevant cases and codes, SIROCCO takes the following approach. Given the foundational cases, the program:

a) Accepts a new fact situation and a question raised by that fact situation (together the fact situation and question are known as the *target case*), transcribed into ETL.

b) Retrieves and matches cases in the case base (i.e., the *source cases*) using a two-stage algorithm. Stage 1 is a fast and efficient, but somewhat coarse, retrieval based on matching, at various levels of abstraction, the Fact Primitives of the target case to the Fact Primitives of all of the source cases. Temporal considerations do not come into play until Stage 2, a more-expensive structural mapping that employs A* search and focuses on the most critical facts and on the chronology of facts.

c) Frames an analysis of the target case by suggesting codes, past cases, and operationalizations that may be relevant to the new case. To harvest its suggestions, the program applies various selection heuristics, strongly influenced by the operationalization techniques, to the results of the retrieval step.

To represent temporal relationships, *Time Qualifiers* are defined between Facts in each Fact Chronology. The ten Time Qualifiers used by SIROCCO are shown in Figure 1. Each Fact in the chronology must have at least one Time Qualifier but may have more than one. Each of the 10 Time Qualifiers represents a disjunctive group of one or more of Allen's temporal relations [1]. The Allen relations provide a formal way to propagate temporal relations throughout a Fact Chronology. Because the case enterer typically does not provide temporal relationships between all Facts in a chronology, it is necessary to compute these relationships when they are not defined. SIROCCO does this by calling a time propagation and management system, TIMELOGIC [12]. TIMELOGIC determines the temporal relationship between any pair of Facts in a chronology through a process of forward chaining over the Allen relations.

---

[1] A fact situation is defined by a single Fact Chronology. A case is defined as a Fact Chronology and a chosen Questioned Fact from that chronology. Thus, if the fact situation raises multiple ethical questions, one fact situation can contain multiple cases.

McLaren, B.M. & Ashley, K.D. (2011). Can temporal representation and reasoning make a difference in automated legal reasoning? Lessons from an AI-based ethical reasoner. In the *Proceedings of the 13th International Conference on Artificial Intelligence and Law (ICAIL-2011)*.

Figure 1 shows how the Time Qualifiers map onto Allen's temporal relations. As an example, consider "After the start of …" from Figure 1. This qualifier specifies that a Fact may: be after (A), during (D), finish (F), be met by (Mi), or be overlapped by another Fact (Oi).

Allen's relations were not used directly, given the intuition that they would not be easily usable by our independent case enterers. For instance, how often does one say that event A "meets" event B? (Allen's language) Rather, an emphasis in defining the Time Qualifiers was to pick temporal relations naturally expressed by humans. Also, using Allen's relations directly would have forced the case enterers to provide disjunctive groups of relations between Facts, a difficult and unnatural task, instead of providing single relations between Facts as with the Time Qualifiers. Of course, abstracting the temporal relations introduces some imprecision. On balance, however, this appeared to be an important and necessary trade-off to achieve the goal of having independent case enterers transcribe the engineering ethics cases.

Finally, we recognize that there are newer technical approaches to representing temporal relations, for instance, Semantic Web processing that accounts for time relations between events (e.g., [9]). However, while the newer Semantic Web representations are aimed at capturing and reasoning about critical web-based events, such as the creation and updating of web pages and access to web services, the essential temporal constraints and relations are not fundamentally different, in a semantic sense, from those of Allen. Thus, while SIROCCO's time qualifiers are based on a long-established temporal model, we argue that the essential meaning of and reasoning about the time qualifiers have not fundamentally changed with newer web-based representations.

SIROCCO's Time Qualifiers	Allen's Temporal Relations
1. Pre-existing fact[2]	B, C, Fi, M, O
2. After the start of ...	A, D, F, Mi, Oi
3. Starts at the same time as ...	S, Si
4. <X time> after the start of ...	D, F, Oi
5. After the conclusion of ...	A, Mi
6. Immediately after the conclusion of ...	Mi
7. <X time> after the conclusion of ...	A
8. Ends ...	F, Mi, Oi
9. Occurs during ... / Occurs as part of ...	D, F, S
10. Occurs concurrently with ...	E

Figure 1: Mapping of SIROCCO's Time Qualifiers to Allen's Temporal Relations

[2] "Pre-existing fact" defines a global time relation that starts before all other Facts in the chronology, except other Facts that are also pre-existing. All of the other qualifiers explicitly designate a Fact or Facts (depicted by ellipses (…) in the table) for which the specified time relation holds.

5. CASE ACQUISITION

Using the case-acquisition web site, twelve independent case enterers represented all of the foundational cases and two of these enterers represented the trial cases.

Time Qualifier : Explanation of Use

After the conclusion of… : Use this qualifier when the only information known is that Event B starts after the conclusion of Event A. If it is also known that Event B starts immediately after the conclusion of Event A, use the more specific qualifier "Immediately after the conclusion of …" If it is also known that Event B starts a specified amount of time after the conclusion of Event A, use the more specific qualifier "[X time] after the conclusion of…"

What is known:

- Event B clearly starts after the conclusion of Event A.

What is not necessarily known:

- Whether Event B starts immediately after the conclusion of Event A.
  Subsumed Qualifier: "Immediately after the conclusion of…"
- Specifically how long after the conclusion of Event A Event B starts.
  Subsumed Qualifier: "[X time] after the conclusion of…".

Ends … : Use this qualifier when Event B signifies, or triggers, the end of Event A. For instance, use it in the following situation:

1. Engineer A **<is employed by>** Company X.
2. Engineer A **<resigns from>** Company X. Ends 1

Typically, the type of "terminating event" represented by 2, above, cannot be specifically defined as occurring before, during, or after the "terminated event." But the duration of the "terminating event" clearly overlaps, in some way, with the end time of the "terminated event."

What is known:

- Event B starts after the start of Event A.
- Event B either finishes, meets, or overlaps with the conclusion of Event A.

What is not necessarily known:

- Whether Event B starts after the conclusion of Event A.
  Overlapping Qualifier: "After the conclusion of…"
  Subsumed Qualifier: "Immediately after the conclusion of…"
- Whether Event B occurs during Event A.
  Overlapping Qualifier: "Occurs during…" / "Occurs as part of…"

After the start of … : Use this qualifier when the only information known is the relative starting time of two events, i.e., that Event B starts after the start of Event A.

What is known:

- Event B clearly starts after the start of Event A.

What is not necessarily known:

- How long after the start of Event A Event B starts.
  Subsumed Qualifier: "[X time] after the start of…"
- Whether Event B starts after the conclusion of Event A.
  Subsumed Qualifier: "After the conclusion of…"
- Whether Event B starts immediately after the conclusion of Event A. *Subsumed Qualifier*: "Immediately after the conclusion of…"
- If Event B starts after the conclusion of Event A, how long after the conclusion of Event A.
  Subsumed Qualifier: "[X time] after the conclusion of…"
- Whether Event B overlaps with the ending of Event A. *Subsumed Qualifier*: "Ends…"

Figure 2: Case enterers' instructions re some Time Qualifiers

The web site contains a Participant's Guide with instructions on how to transcribe ethics cases into EETL, a reference shelf of useful materials, including the full vocabulary of EETL, and an example set of 47 transcribed fact situations. The case enterers submitted each case as a structured set of tables; those tables are subsequently translated into SIROCCO knowledge structures by a PERL program.

For the case enterers, the Time Qualifiers are not described in terms of Allen's relations, but, instead, in terms that are easier to understand and more relevant to the transcription task. In particular, each qualifier has associated information that is intended to guide the case enterer's choice – for example its intended use, what one needs to know to apply it, and links to other possible qualifiers and to case examples.

Examples of some of the Time Qualifiers and an explanation of how each is used are provided in Figure 2. These explanations are slightly edited versions of those provided to the case enterers on the case-acquisition web site. Whenever "Event" is mentioned in the table, this could refer to either an action or an event, i.e., a Fact in a Fact Chronology.

6. SIROCCO STAGE 2 EXAMPLE

As noted in the previous section, in Stage 1 the program retrieves a set of source cases that match at least one of the target's Fact Primitives, or an abstraction thereof. The cases are assigned scores based on the degrees to which their Fact Primitives overlap the target's and their critical and questioned facts match the target's Facts. The Stage 1 scoring formula, however, does not take time order into account; Stage 2 does.

After scoring all of the source cases, Stage 1 passes the top N cases to Stage 2 for structural mapping. The structure of a case is a graph that is defined by (1) the individual Facts' Actors, Objects, Fact Primitives, and combinations thereof (i.e., Fact Phrases) and (2) the temporal relationships between Facts represented by the Time Qualifiers and the propagation of those qualifiers. The program performs a graph mapping of the target case to each of the N cases, using A* heuristic search focusing only on critical facts as indicated in the Code Instantiations and Case Instantiations found in the top-rated source cases. As noted, these instantiations, a type of operationalization, relate the questioned fact, certain critical Facts, and the temporal sequence of those Facts to a code or case citation. The structural mappings are scored and sorted according to percentage of match and are passed to the next phase of the program which selects and displays the final output of the program (i.e., suggested codes, cases, and operationalizations). In scoring, SIROCCO tests whether the mapped Facts in the source have an analogous temporal relation to the corresponding Facts in the target. A pair of Facts in the source is temporally analogous to a pair of Facts in the target if the temporal relationship of the source's paired facts *intersects* with the temporal relationship of the target's pair in terms of Allen's temporal relations. A set of Facts in the source is temporally analogous to a set of Facts in the target if every pair of facts in the source is temporally analogous to the corresponding pair of facts in the target.

In section 3, we saw an example of a temporal distinction between Case 62-7-1 and target Case 91-6-1 in relation to Code III.1.c. Here is how SIROCCO can identify this difference between the cases. First, consider a version of the program that *does not* take temporal considerations into account. Figure 3 shows a structural mapping that does not impose temporal constraints, generated in this so-called NON-TEMP SIROCCO's Stage 2, between *Code Instantiation* III.1.c. of source Case 62-7-1 and target Case 91-6-1.

Source Facts (Case 62-7-1)	**Mapped Target Facts (Case 91-6-1)**
C1-2. Engineer Z <**is employed by**> County Metropolitan Commission <as> Engineer & Advisor.	C2-1. Federal Environmental Agency <**hires the services of**> Engineer A <**for**> Hazardous Waste.
C1-3. Engineer Z <**is hired to provide services for**> Developer D.	C2-6. Major Industrial Corporation <**hires the services of**> Engineer A <**for**> Hazardous Waste.
C1-4. Engineer Z <**in his capacity as**> Advisor <**takes the action**> (Engineer Z <**negotiates with**> Developer D <**for**> Housing Development H).	C2-8. Engineer A <**provides engineering services on**> Hazardous Waste <**for**> Major Industrial Corporation.

Figure 3: Structural mapping between two cases

In this structural mapping the key mappings between the source and the target are the facts that indicate that each engineer worked for two separate clients (i.e., Client1: C1-2 ←→ C2-1 and Client2: C1-3 ←→ C2-6). If temporal considerations are disregarded the above table represents a valid mapping from the source to the target. Because of the resulting favorable match score of this Code Instantiation, the NON-TEMP version of SIROCCO that does not take temporal considerations into account would ultimately, but incorrectly, propose III.1.c. as a possibly relevant code when advising the user about target case 91-6-1.

In a mapping of the same Instantiation, however, generated by a version of SIROCCO that *does* take temporal considerations into account, the cross-case mappings between facts C1-3 and C2-6 and between facts C1-4 and C2-8 survive, but the cross-case mapping between C1-2 and C2-1 fails. The temporal relations between Facts C1-2 and C1-3 of the source do not overlap with the temporal relationships between corresponding Facts C2-1 and C2-6 of the target. More specifically, Facts C1-2 and C1-3 of the source have a Time Qualifier of "occurs during" (5 in Figure 1), while Facts C2-1 and C2-6 of the target have a Time Qualifier of "after the conclusion of" (9 in Figure 1). Because these pairs of Facts are not temporally analogous, the node representing a structural mapping of all of these Facts is rejected as invalid. Because this relatively poorer mapping is assigned a relatively poor match score, the Instantiation is ultimately rejected and Code III.1.c. is not proposed as a possibly relevant code. This action corresponds to what the Board actually did. Since the analysis of the NSPE BER cases turned up a number of instances in which the Board applied similar temporal considerations, we proposed a temporal hypothesis.

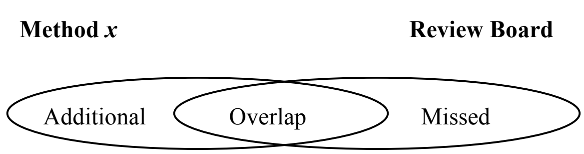
7. EXPERIMENT TO ASSESS TEMPORAL KNOWLEDGE IN SIROCCO

Our temporal hypothesis is:

Temporal knowledge is integral in the retrieval and analysis of principles and past cases. A computational model *incorporating temporal knowledge will improve the accuracy of the model's predictions and analysis.*

To test the temporal hypothesis, a set of 58 *trial cases* was provided as input both to an ablated version of SIROCCO, NON-TEMP SIROCCO, that *does not* employ temporal knowledge, and to SIROCCO itself, the regular version of the program that *does* employ temporal knowledge. All of the trial cases were transcribed into ETL by two independent case enterers and the resulting case representations, unlike the foundational cases, were provided unaltered to SIROCCO for processing to ensure objectivity. The 58 trial cases were chosen from two pools of cases within a set of 77 cases decided by the NSPE BER between 1993 and 1998: 44 trial cases were chosen at random from 52 selected topics cases and 14 trial cases were chosen at random from 25 non-selected topics cases. For both SIROCCO and NON-TEMP SIROCCO, the aforementioned 184 foundational cases were used as the case base for retrieval of past cases and codes.

The results of NON-TEMP SIROCCO were compared against the suggestions made by the ethical review board and an F-Measure calculated for each individual sample and as a mean value over all samples. These results were then compared to the output of the standard version of SIROCCO, which *did* apply temporal knowledge.



Method *x* = The code and case suggestions produced by SIROCCO or one of the other methods.

Review Board = The code and case citations of the ethics review board.

Overlap = The code and case citations shared by both Method *x* and the review board.

Additional = The extra code and case suggestions made by Method *x* but not cited by the review board.

Missed = The set of code and case suggestions lacking in Method *x* but cited by the review board.

Precision (*P*) = Overlap / (Overlap + Additional)

Recall (*R*) = Overlap / (Overlap + Missed)

F-Measure[3] (Method *x*, Review Board) =

$$\frac{(\beta^2 + 1)PR}{\beta^2 P + R}$$

Figure 4: Calculation of the F-Measure in the SIROCCO Experiments

The *F-Measure* [24, p. 173-176; 13] is a metric that combines precision and recall. Informally, the F-Measure was used to compute a form of overlap of the codes and cases between two methods. A Venn diagram depicting the overlap and the equations for precision, recall, and the F-Measure is shown in Figure 4.

Two F-Measures were calculated for each method for each case, one representing combined exact matches of codes and cases between the method's solution and the Board's, and one representing combined inexact matches of codes and cases. Codes and cases were combined to simplify the comparisons between methods. The combined exact F-Measure was calculated by comparing codes and cases separately but treating them as being in one group for purposes of calculating precision and recall. In other words, the corresponding numerators and denominators for the precision and recall calculations of the code and case comparisons were added together. For instance, if Method *x* suggested Code A, Code B, and Case X, and the review board cited Code A and Case X for the same case, then, assuming ß = 1.0, the F-Measure calculation would be:

P (Method x, Review Board) = (1 + 1) / (2 + 1) = 0.67

R (Method x, Review Board) = (1 + 1) / (1 + 1) = 1.0

F-Measure (Method x, Review Board) = (2 * 0.67 * 1.0) / ((1 * 0.67) + 1.0) = 0.8

The combined inexact F-Measure required different treatment of the codes and cases. Codes were considered matched if they resided in the same category of the Code Hierarchy. Exact code matches were first identified, tallied, and removed. The remaining codes in the "Additional" and "Missed" categories (see the Venn diagram of Figure 4) were then pairwise compared to check for matches at higher abstraction levels of the Code Hierarchy. Any abstractly matching pairs were considered "equal" for the purposes of the inexact F-Measure calculation. An abstractly matched pair is any pair that shares the same top-level abstraction category. There are a total of 22 such categories.

Calculating the case-matching component of the combined inexact F-Measure was more complicated. Cases were not grouped by a predefined knowledge structure such as the Code Hierarchy, and so there were no static means of computing equivalence or relevance between pairs of cases. Yet it was also true that if the Board did not cite a case, say Case A, this did not necessarily imply that Case A was irrelevant. Rather, it might indicate that the Board was unable, or perhaps did not have the time, to find such a case for their analysis. Because there are significantly more cases to choose from than codes, it is far less likely that the Board will cite a given case than a given code. Alternatively, the Board might have simply believed they had made themselves clear without citing the case. The Board cited, on average, significantly fewer cases than codes per case. For the 475 cases decided between 1958 and 1992, the Board averaged 2.24 code citations per case and 1.44 case citations per case. It was not unusual, in fact, for the Board to cite no cases in their analyses – this happened in 40% (73 out of 184) of the foundational case analyses – while only very few of the Board's case analyses had no code citations.

It was of course necessary to find a fair way of assigning credit to each method for finding cases that seemed to be (or are) relevant, even if they did not show up in the Board's opinions. In other words, credit was required to be given for a relevant case suggestion, even if there was no corresponding case cited by the

[3] ß < 1.0 gives greater weight to precision; ß = 1.0 gives equal weight to recall and precision; ß > 1.0 gives greater weight to recall. ß = 1.0 was used for all experiments reported in this dissertation.

target case. With respect to Figure 4, this notion was quantified as follows: For each case suggestion made by Method *x*, credit was given for an inexact match to either (a) the target case or (b) one of the cases the Board cited in its analysis of the target. There might be multiple matches to the target, but each cited case could match only once. For the purposes of the F-Measure calculation, therefore, each inexact match increased "Overlap" by 1 and decreased "Additional" and "Missed" by 1. When "Missed" reached 0, it was no longer decremented. Notice that in the relatively common circumstance of the Board citing no cases, the calculation amounted to simply incrementing "Overlap" by 1 for each inexact match. If a target case cited at least one case, the "Overlap" for the recall calculation would have a maximum value equal to the number of cases cited by the target.

The altered version of the F-Measure metric used to handle inexact matching – particularly that of case citations – is a reasonable and fair metric with which to benchmark the computational methods. While the precise definitions of precision and recall were slightly altered, the spirit and intent of the metrics remain intact. In particular, the altered versions of the F-Measure, precision, and recall capture the "correctness" of a method's selections with respect to both the Board's citations and additional knowledge sources (i.e., the Code Hierarchy and the citation overlap metric).

To calculate the combined inexact F-Measure, the corresponding numerators and denominators of code and case precision and recall were added, as with the exact calculation. As an example, suppose Method *x* suggested Code A, Code B, Case X, and Case Y, while the review board, for the same case, cited Code C and Case Z. If Code A and Code C inexactly matched, Case X inexactly matched the target case, but Case Y did not inexactly match either Case Z or the target, then the F-Measure calculation would be:

P (Method x, Review Board) = (1 + 1) / (2 + 2) = 0.5

R (Method x, Review Board) = (1 + 1) / (1 + 1) = 1.0

F-Measure (Method x, Review Board) = (2 * 0.5 * 1.0) / ((1 * 0.5) + 1.0) = 0.67

8. RESULTS

The results of running SIROCCO and NON-TEMP SIROCCO on all of the trial cases were as follows. For SIROCCO, the mean F-Measures were 0.212 for exact matching and 0.462 for inexact matching. For NON-TEMP SIROCCO the mean F-Measures were 0.213 for exact matching and 0.461 for inexact matching. Not surprisingly, and as shown in Table 1, these results indicate that SIROCCO was *not* significantly more accurate than NON-TEMP SIROCCO with respect to the trial cases. In fact, the results were very much the same for both versions of the program. This result can be visualized by inspecting the paired differences charts of Figure 5. For both the exact and inexact matching criteria, a very high percentage of the F-Measures calculated by SIROCCO and NON-TEMP SIROCCO for the same cases were identical, as witnessed by the large bar at the 0 hash mark of both charts. Only a small percentage of the cases led to different F-Measures. Most of these are found close to the 0 hash mark, meaning that the absolute differences were slight.

Table 1: Comparison of SIROCCO and NON-TEMP SIROCCO using Nonparametric Bootstrapping

Methods Compared	**95% Confidence Interval for mean difference**	**p-value for mean difference=0**
SIROCCO vs. NON-TEMP SIROCCO, Exact Matching	(-0.017, 0.011)	0.909
SIROCCO vs. NON-TEMP SIROCCO, Inexact Matching	(-0.021, 0.025)	0.926

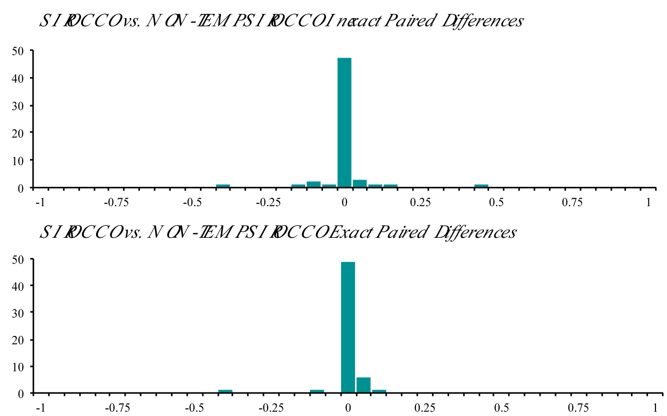


Figure 5: Experiment #3, Paired Differences of the F-Measures for Exact and Inexact Citations

A similar, informal test was also run with the foundational cases. In this test, the mean F-Measures for exact matching were 0.285 for SIROCCO and 0.292 for NON-TEMP SIROCCO; the mean F-Measures for inexact matching were 0.479 for SIROCCO and 0.473 for NON-TEMP SIROCCO. Although the p-value for a mean difference = 0 for exact matching was slightly more favorable on this data set (i.e., 0.148), it was still far less than significant. In addition, the p-value for inexact matching was not significant, as shown by the p-values in Table 1. Thus, this test tended to confirm the results of the experiment – that SIROCCO's temporal knowledge did not improve retrieval accuracy.

9. DISCUSSION

It is clear from the results of the experiment, as well as from the informal follow-up test with the foundational cases, that SIROCCO's application of temporal knowledge did not make a difference in the overall accuracy of code and case retrieval. In other words, the temporal hypothesis was not supported by the results of this experiment.

Why didn't SIROCCO's temporal knowledge make a statistical difference? As discussed and illustrated in section 6 there clearly are specific instances in which SIROCCO's temporal knowledge can be used to identify important distinctions between cases and thereby lead to the correct selection of relevant codes and cases.

However, the experiment shows that while specific instances may benefit from the application of temporal knowledge, accuracy in general is not improved by SIROCCO's current implementation and use of temporal knowledge.

One possible reason for this is evident in the charts found in Figure 5. As discussed above, a significant majority of the trial cases exhibited no difference in the F-Measures attained by SIROCCO versus those attained by NON-TEMP SIROCCO. This suggests that a large number of the trial cases simply did not involve temporal considerations, or at least did not involve them in a critical fashion. A cursory reading of a subset (10) of the trial cases indicated that this may be generally true. While most of the 10 cases involved temporal events, only a couple appeared to turn specifically on event sequence. In most of the cases, the mere existence of certain facts was more critical than the order of those facts.

However, such an explanation cannot fully explain why temporal knowledge also did not play a significant role in the test executed with the foundational cases. As discussed in sections 3 and 6, the analysis of the NSPE BER cases seemed to indicate that temporal knowledge *did* play a role in the decision making of the review board and many of those cases were included in the foundational case base. On the other hand, even though a fair number of the foundational cases did rely on temporal considerations, a larger number did *not* rely on temporal considerations.

There are other explanations for this result. It might be relatively rare that pairs of cases exist such that (1) a difference in temporal ordering leads to two different ethical interpretations and (2) both cases are found in the NSPE BER case base. In other words, SIROCCO's opportunities to identify important differences in cases, based on temporal differences, may be rare. Thus, it may be that the only way to truly test SIROCCO's capability to discern differences in cases based on temporal considerations is to modify hypothetically cases such that the time ordering changes and a case that once involved an ethical obligation no longer does. Running SIROCCO on a test set of such cases could test whether the program successfully distinguishes among the meaningful and unmeaningful scenarios.

On the other hand, some facts indicate the relevance of specific codes regardless of temporal considerations and would thus be unaffected by any differences in temporal ordering. For instance, consider example case 90-5-1, a case in which tenants of an apartment building sued the owner to force him to repair many defects in the building that affected the quality of use. The owner's attorney hired an engineer to inspect the building and give expert advice. The engineer discovered serious structural defects in the building that posed an immediate threat to the safety of the tenants and that were not mentioned in the tenants' complaint. Upon reporting the findings to the attorney, the attorney told the engineer to maintain this information as confidential as it was part of a lawsuit, and the engineer complied. The fact that the tenants' safety in the apartment building was at risk indicates the relevance of a public safety code. While deciding whether such a code is violated may depend on the sequence of temporal events, the relevance of the code probably requires less specific knowledge of the order of events.

Another possible explanation for the results of the experiment is that certain Facts typically occur in approximately the same temporal position in Fact Chronologies, thus producing less variability and perhaps less temporal distinction between the chronologies. For instance, it certainly only makes sense that one would report a dangerous situation after one learns of it, not before, so any cases involving such an issue would naturally *always* incorporate this particular sequence of events. As another example, primitives dealing with employment, such as "hires the services of" or "is employed by," tend to be at the beginning of Fact Chronologies and, in addition, are quite often assigned the Time Qualifier "Pre-existing fact." Such a Fact designated as a Critical Fact of a *Code Instantiation* or *Case Instantiation* is less likely to produce a differentiating effect in the structural mapping process.

Case enterers' inaccurately assigning Time Qualifiers is another possible explanation. As discussed in sec. 4 the task of assigning temporal relations between Facts is not easy for people and although the Time Qualifiers– disjunctive groups of Allen's temporal relations – are intended to somewhat ease this difficulty, the task is still hard and prone to error. In fact, because of obvious human error, some post editing of the foundational cases provided by independent case enterers was necessary and involved the correction of Time Qualifiers.

Finally, the particular implementation and use of temporal knowledge within SIROCCO may be a contributing factor. First, the underlying use of disjunctive groups of Allen's temporal relations between Facts tends to lead to a somewhat "forgiving" approach in accepting structural mappings. Because SIROCCO's algorithm performs an intersection of the temporal relations between pairs of Facts, there may be situations in which only a single, common possible relation allows a mapping to succeed. While this is strictly correct, it may lead to the acceptance of unlikely mappings. Second, temporal knowledge may simply not have a big enough role in SIROCCO's structural mapping scoring algorithm and selection heuristics. For example, a critical subsequence of steps in a *Code Instantiation* may be rejected according to SIROCCO's algorithm, but the *Code Instantiation* overall may still succeed if other Facts match well enough for the Instantiation to receive a relatively high score. Also, the selection of codes and cases by SIROCCO's Analyzer is based on an *accumulation* of evidence supporting individual codes and cases and not on the success of individual mappings. This implies that the failure (or success) of individual mappings may not always be enough to alter the suggestions made by the program. A final implementation issue is that all temporal relations are treated equally when, in fact, certain temporal relations might be critical, while others are not. If the case enterer had a means of designating *critical* temporal relations, and SIROCCO's algorithm was sensitive to these, this might make a difference in SIROCCO's accuracy.

Most likely, some combination of the above reasons underlies SIROCCO's failure to gainfully use temporal knowledge in the selection of codes and cases. Although the results of this experiment do not support the temporal hypothesis, it is possible that addressing some of the case-acquisition and implementation issues discussed above might have made a difference. This difference might occur if a significant number of the cases do, in fact, rely on temporal considerations, as was conjectured from the analysis of the NSPE BER cases.

10. RELATED WORK

Case-based planning systems, such as CAPLAN/CBC [21], PRODIGY/ANALOGY [26], and CHEF [8] involved temporal sequencing in the solutions they generated, stored and retrieved, but the retrieval indices of such systems consisted of elements such as goals, partial ordering between the goals, and initial state descriptions, not the temporal relations between individual steps, as in SIROCCO. The solution sought by a planning system is a set of temporally related steps, while the solution sought be SIROCCO is a *normative analysis* of a *given* set of temporally related steps.

Gardner's and Yoshino's systems performed such normative (legal) analyses of problems involving contract law; given the nature of the law of offer and acceptance, these programs perforce took into account the temporal ordering of events [6; 27].

By contrast, SIROCCO performs a case-based (and code-based) normative analysis; thus, it retrieves cases (and code provisions) that are similar based on their temporal, as well as other, features. BROADWAY, a world-wide web browsing advisor [11], used temporally related steps to assist in retrieval. It used a current sequence of browsed pages as indices into a case base of previous browsing sessions, represented as time-extended situations, to advise a user on what to do next. This system used much simpler temporal constraints (i.e., before, after) than SIROCCO, but a more general architecture that performed matching of the Allen primitives underlies the application [10].

Subsequent to SIROCCO, Ma and Knight [14] presented a general framework for retrieving case histories, for example in the medical domain, with a representation of relative and absolute temporal knowledge. The framework employs "fluents", propositions whose truth values depend on time, and "elemental cases", time-independent episodes defined as collections of fluents. Case histories are represented as lists of elemental cases with temporal relationships specified via time elements, a "meets" relation over time elements, and a "holds" relation of fluents over time elements. Non-temporal similarity involves elemental case matching, while temporal similarity is assessed using conventional graph similarity measurement of graphs representing temporal information as meets relations and durations. Montani and Portinale [20] developed a method for retrieving medical cases with similar time series.

11. CONCLUSION

Since we were unable to confirm the temporal hypothesis, with the program, a number of changes relevant to temporal reasoning could be explored. For instance, currently SIROCCO treats all instances of Time Qualifiers as equal. However, just as particular Facts of a chronology are more important than others, it is also true that particular Time Qualifiers are more important than others. Thus, a scheme in which Time Qualifiers can be designated as more crucial, leading to greater or lesser match scores, might be helpful. Another possible change related to weighting the Time Qualifiers would involve the development of a capability to override SIROCCO's default approach of accumulating evidence in support of particular codes and cases. In particular, there may be instances in which a temporal match (or mismatch) completely outweighs the combined evidence from a variety of other case or code matches. Thus, the development of an exception handler, or specialized heuristics, within SIROCCO's Analyzer might be fruitful.

The experiment also suggests the need to explore more analytically the likelihood that temporal orderings matter in distinguishing among cases in particular datasets. Walton [28] and Ashley [2] were discussing the need to distinguish among superficially similar cases involving different legal claims where temporal orderings would reveal underlying differences. While it seemed likely that the NSPE cases, dealing with many different ethics code provisions, would be similar to legal datasets involving multiple claim scenarios, we could not confirm empirically that temporal orderings did in fact matter.

Apart from the experiment, to the extent that narrative representations of cases become important in AI and Law, SIROCCO's approach to representing and reasoning has much to recommend it. The program supports a language for representing the temporal ordering of events in ethics cases in a way that informed determinations of whether and how ethical norms were violated and if the problem was normatively analogous to other cases. Its measure of temporal analogy goes beyond surface features to determine if cases are relevantly similar at a deeper level. At the same time, the program supports ordinary case enterers in translating the facts of textually described cases into a representation that the program can process. While SIROCCO deals with engineering ethics cases, it is clear, at least in theory, that similar temporal considerations apply in legal cases and that the same Time Qualifiers should be useful in legal narratives.

12. REFERENCES

[1] Allen, J.F. (1983) Maintaining Knowledge about Temporal Intervals. In the *Communications of the ACM* 26(11), 832-843.

[2] Ashley, K.D. (2009) Ontological Requirements for analogical, teleological and hypothetical reasoning. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law* (ICAIL-09) pp. 1-10. ACM Press.

[3] Ashley, K.D. and McLaren, B.M. (2001). An AI Investigation of Citation's Epistemological Role. In the *Proceedings of the Eighth International Conference on Artificial Intelligence and Law* (ICAIL-01). pp. 32-41. ACM Press.

[4] Berman, D.H. and Hafner, C.D. (1995) Understanding Precedents in a Temporal Context of Evolving Legal Doctrine. In the *Proceedings of the Fifth International Conference on Artificial Intelligence and Law* (ICAIL-95) pp. 42-51. ACM Press.

[5] Bex, F. (2009). Evidence for a Good Story: A Hybrid Theory of Arguments, Stories and Criminal Evidence. PhD Thesis: University of Groningen.

[6] Gardner, A. vdL. (1987) *An Artificial Intelligence Approach to Legal Reasoning*. MIT Press, Cambridge.

[7] Governatori, G., Rotolo, A. Riveret, R., Palmirani, M. and Sartor, G. (2007) Variants of Temporal Defeasible Logics for Modelling Norm Modifications In the *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law* (ICAIL-07) pp. 155-160. ACM Press.

[8] Hammond, K.J. (1989) *Case-Based Planning: Viewing Planning as a MemoryTask*. San Diego, CA: Academic Press, Inc.

[9] Hobbs, J.R. and Pan, F. (2004). An Ontology of Time for the Semantic Web. ACM Transactions on Asian Language Information Processing (TALIP) - Special Issue on Temporal Information Processing TALIP Homepage archive Volume 3 Issue 1, March 2004

[10] Jaczynski, M. (1997) Framework for the Management of Past Experiences with Time-Extended Situations. In the *Proceedings of the Sixth ACM Conference on Information and Knowledge Management* (CIKM '97), Las Vegas, Nevada.

[11] Jaczynski, M. and Trousse, B. (1998) BROADWAY: A World Wide Web Browsing Advisor Reusing Past Navigations from a Group of Users. In the *Proceedings of the Fourth European Workshop on Case-Based Reasoning*. Dublin, Ireland.

[12] Koomen, J.A. 1989. *The TIMELOGIC Temporal Reasoning System*. Tech. Rep. 231, Computer Science Dept., U. Rochester, NY.

[13] Lewis, D.D., Schapire, R.E., Callan, J.P. and Papka, R. (1996). Training Algorithms for Linear Text Classifiers. In the *Proceedings of the 19th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*. Zurich, Switzerland, 1996.

[14] Ma, J. and Knight, B. (2003) A Framework for Historical Case-Based Reasoning. In *Case-Based Reasoning Research and Development*; Lecture Notes in Computer Science, Vol. 2689/2003, p. 1067. Springer.

[15] McLaren, B.M. (1999) Assessing the Relevance of Cases and Principles Using Operationalization Techniques. Unpublished Ph. D. Dissertation, Graduate Program in Intelligent Systems, University of Pittsburgh.

[16] McLaren, B.M. (2003). Extensionally defining principles and cases in ethics: An AI model. *Artificial Intelligence Journal*, *150*, 145-181.

[17] McLaren, B.M. and Ashley, K.D. (1998). "Exploring the Dialectic between Abstract Rules and Concrete Facts". In *Advances in Case-Based Reasoning. Proceedings, Fourth European Workshop on Case-Based Reasoning* (EWCBR-98) Dublin, Ireland. B. Smyth and P. Cunningham (eds) pp. 37-51. Lecture Notes in Artificial Intelligence 1488. Springer: Berlin.

[18] McLaren, B.M. and Ashley, K.D. (2000). "Assessing Relevance With Extensionally Defined Principles and Cases". In Proceedings of the 17th National Conference of Artificial Intelligence (AAAI-00). pp. 316-322. Austin. August. AAAI Press, Menlo Park, CA.

[19] McLaren, B.M. and Ashley, K.D. (2001) "Helping a CBR Program Know What it Knows". In *Proceedings of the Fourth International Conference on Case-Based Reasoning.* (ICCBR-01) *Case-Based Reasoning Research and Development*, (Aha, D. and Watson, I. eds.) Lecture Notes in AI Nr. 2080. pp. 377-391. Springer.

[20] Montani, S. and Portinale, L. (2005) Case Based Representation and Retrieval with Time Dependent Features. In *Case-Based Reasoning Research and Development.* Lecture Notes in Computer Science, Vol. 3620/2005, pp. 353-367, Springer.

[21] Muñoz-Avila, H. and Huellen, J. (1995) Retrieving Cases in Structured Domains by Using Goal Dependencies. In the *Proceedings of the First International Conference on Case-Based Reasoning*, pp. 241-252. Sesimbra, Portugal.

[22] NSPE (1958-1998), National Society of Professional Engineers. Opinions of the Board of Ethical Review, Volumes I through VIII. Alexandria, VA: the National Society of Professional Engineers, 1958-1998.

[23] Palmirani, M., Governatori, G. and Contissa, G. (2010) Temporal Dimensions in the Rules: an Evolution of LKIF Rule. In *Proceedings of the 23rd International Conference on Legal Knowledge and Information Systems*. pp. 159-162. IOS Press.

[24] van Rijsbergen, C.J. (1979). C.J. van Rijsbergen. *Information Retrieval*. London: Butterworths, Second edition.

[25] Rissland, E.L. and Friedman, M.T. (1995) Detecting Change in Legal Concepts in *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*. pp. 127-136. Association for Computing Machinery. New York. May.

[26] Veloso, M. (1992) *Learning by Analogical Reasoning in General Problem Solving*. Ph.D. Dissertation, Carnegie Mellon University.

[27] Vila, L. and Yoshino, H. (2005) Time in Automated Legal Reasoning. In Fisher, M., Gabbay, D., and Vila, L. (eds.) Handbook of Temporal Reasoning in Artificial Intelligence. pp. 537-558. Elsevier: Amsterdam.

[28] Walton, D. (2010) Similarity, Precedent and Argument from Analogy. *Artificial Intelligence and Law* pp. 217-246.

[29] Wyner, A. and Peters, W. (2010) Lexical Semantics and Expert Legal Knowledge towards the Identification of Legal Case Factors. In *Proceedings of the 23rd International Conference on Legal Knowledge and Information Systems.* pp. 127-136. IOS Press.