Stefanie Brüninghaus and Kevin D. Ashley (2001) Improving the Representation of Legal Case Texts with Information Extraction Methods.
In: Henry Prakken and Ron Loui (Editors) Proceedings of the Seventh International Conference on Artificial Intelligence and Law
(ICAIL-2001). St. Louis, MO. ACM Press: New York, NY. Pages

Improving the Representation of Legal Case Texts with Information Extraction Methods

Stefanie Brüninghaus and Kevin D. Ashley
Learning Research and Development Center,
Intelligent Systems Program and School of Law
University of Pittsburgh
Pittsburgh, PA 15260
steffi+@pitt.edu, ashley+@pitt.edu

Abstract

The prohibitive cost of assigning indices to textual cases is a major obstacle for the practical use of AI and Law systems supporting reasoning and arguing with cases. While considerable progress has been made toward extracting certain facts from well-structured case texts or classifying case abstracts under Key Number concepts, these methods still do not suffice for the complexity of indexing concepts in CBR systems.

In this paper, we show how a better example representation can facilitate classification-based indexing. We present the hypothesis that (1) abstracting from the individual actors and events in case, (2) capturing actions in multi-word features, and (3) recognizing negation, can lead to a better representation of legal case texts for automatic indexing. We discuss how a state-of-the-art information extraction program can be used to implement these techniques. Preliminary experimental results show that a combination of domain-specific knowledge and IE techniques can be used to generalize from the examples and derive more powerful features.

1 Motivation: Indexing and Information Extraction for Legal Cases

AI and Law research has brought forth numerous sophisticated models and formalizations for reasoning and arguing with cases (Ashley 1990; Aleven 1997; Rissland, Skalak, & Friedman 1996; Branting 1999); see also (Prakken & Sartor 1998). However, despite their potential benefits, none of these approaches has lead to systems used in legal practice, for instance as intelligent assistants that help in exploring case law or in generating arguments with cases.

One of the major obstacles for the use of those models in the law offices is the prohibitive cost of representing cases in a form that allows an AI program to reason with them. Up to now, case indexing and information extraction for case-based reasoning (CBR) has been a manual effort. Unless there are methods to facilitate the development and maintenance of large casebases, the scope of these systems will remain restricted to the research world.

The Solomon (Moens 2000) and Prudentia (Weber 1999) projects demonstrated how information can be extracted automatically from civil law criminal cases and used to perform some basic case retrieval. Where the cases are as well-structured as in these applications, key information can be easily spotted, using for instance location and cue-words. However, for many CBR domains, like trade secret law, case structure differs significantly across cases, and these techniques are not applicable. Furthermore, the information extracted in both systems was more concrete and easier to find than the abstract indexing concepts of CBR systems.

Our SMILE[1] project explores new methods to extract information from and index textual cases. It uses a set of marked-up case summaries as examples in a machine learning approach to classify new cases under indexing concepts. In this paper, we illustrate how a better example representation can lead to a better classification under indexing concepts. We discuss the hypothesis that (1) abstracting from the individual actors and events of a case to generic roles, (2) capturing actions in multi-word features, and (3) determining the presence and scope of negation leads to a more powerful representation that allows a classifier to generalize accurately from the training examples.

SMILE goes beyond other approaches in AI and Law because it integrates both domain-specific knowledge and state-of-the-art NLP tools. The integration of background knowledge is the most important difference from the SPIRE system (Daniels & Rissland 1997). Like SMILE's, SPIRE's goal was to overcome the knowledge-engineering bottleneck of indexing cases for a HYPO-style CBR system. SPIRE successfully used a small collection of marked-up text segments to help direct a human indexer to relevant text passages in long cases. For this, SPIRE very elegantly used the relevance feedback and passage retrieval mechanisms of an existing information retrieval (IR) system, which did not allow the integration of any background knowledge or linguistic information.

The rest of this paper is organized as follows. In Section 2, we show an example case and introduce our case-based argumentation program CATO. In Section 3, we introduce the learning approach for assigning indices in SMILE. In Section 4, we discuss why the three techniques for an improved case representation lead to better indexing for legal cases. In Section 5, we demonstrate how to use NLP methods and background knowledge to implement these measures, and present initial experiments with our methods. In Section 6, we summarize the most important issues raised in this paper.

2 Reasoning with Factors in CATO

In this paper, we illustrate our ideas with a hypothetical example, the *Steffi v. Vincent* case. This case is comparable

[1] SMart Index LEarner

to real case summaries, or squibs. In particular, all of the techniques presented here are applicable to the squibs and opinions of real cases; they are not limited to this hypothetical. However, we have not found a good example where all techniques can be applied so clearly in one single case. This said, consider our hypothetical trade secret problem, the *Steffi v. Vincent* case:

> Steffi has developed a method for indexing text cases in her PhD research. *F15 (π): The method is unique* in that it integrates a machine learning approach with background knowledge. *F1 (δ): Steffi tells the method to her friend Vincent.* She asks whether she can test it on his data, and suggests they can write a research paper about the results. *F21 (π): Vincent tells Steffi that he will treat her method as confidential.* Because they are friends, *$\neg$F4 (δ): Vincent does not sign a written agreement.* Vincent, however, uses Steffi's indexing method for his successful educational software start-up company. When Steffi finds out about this, she wants to sue Vincent for trade secret misappropriation.

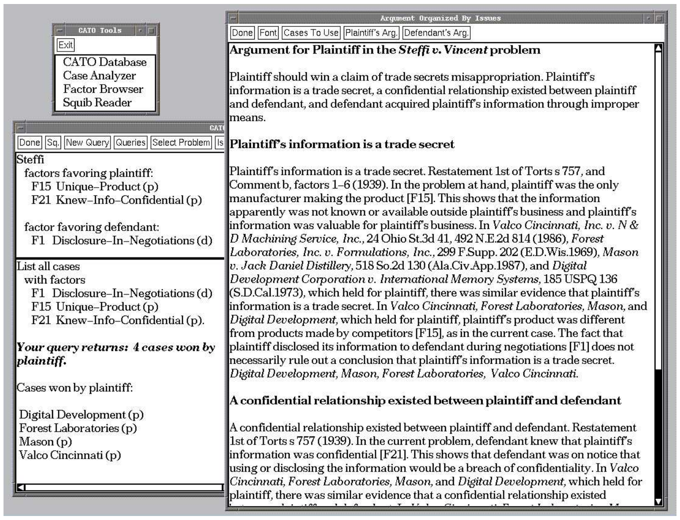
In this case, plaintiff's claim is strengthened to the degree that she had a unique idea, and that the defendant knew the information was confidential. The corresponding sentences are italicized and marked with a π and the relevant pro-plaintiff Factor from the CATO program. Defendant's position is supported by the absence of a non-disclosure agreement and by plaintiff's voluntary disclosure of the information, indicated with δ and the respective Factors.

In a written argument for the plaintiff, her lawyer would address the issues whether the information qualifies as a trade secret, whether there was a confidential relationship, and whether defendant misappropriated the information. For each issue, he would cite favorable cases that share relevant Factors or generalizations with the problem.

A model of this case-based argumentation is implemented in CATO. CATO is an intelligent tutoring environment for teaching skills of making arguments with cases to law students (Aleven 1997). It has an expert model of case-based argumentation, in which it compares cases in terms of 26 prototypical fact situations, the Factors. CATO's Factors are binary features (present in a case or not), and favor either plaintiff or defendant. The system can dynamically generate arguments favoring either side's claim by analogizing or distinguishing precedents. Knowledge of more high-level issues and legal concerns is represented in CATO's Factor Hierarchy, which is used to organize arguments by issues, the way an attorney does, and to reason with partially matched cases and with cases that raise relevant issues (Ashley & Aleven 1997). The Case Database comprises 150 trade secret cases with squibs, and the full-text opinions.

Figure 1 shows how CATO could be used in the*Steffi v. Vincent* problem. In the upper left, the case is expressed in terms of CATO's Factors. The large window in the front shows CATO's argument for the plaintiff, which uses cases retrieved by a query displayed in the lower left.

In a formative evaluation in the context of a legal writing class, CATO's performance compared well with one of the most accomplished instructors for the class. Written exams indicated that students using CATO had learned basic skills of using cases in arguments as well as those students that received small-group classroom instruction, which is considered a very high standard. When students' answers to short argumentation questions were graded, arguments generated by CATO and disguised as a student's solution actually received one of the highest grades.



Steffi
factors favoring plaintiff:
F15 Unique–Product (p)
F21 Knew–Info–Confidential (p)

factor favoring defendant:
F1 Disclosure–In–Negotiations (d)

List all cases
with factors
F1 Disclosure–In–Negotiations (d)
F15 Unique–Product (p)
F21 Knew–Info–Confidential (p).

Your query returns: 4 cases won by plaintiff.

Cases won by plaintiff:

Digital Development (p)
Forest Laboratories (p)
Mason (p)
Valco Cincinnati (p)

Argument for Plaintiff in the *Steffi v. Vincent* problem

Plaintiff should win a claim of trade secrets misappropriation. Plaintiff's information is a trade secret, a confidential relationship existed between plaintiff and defendant, and defendant acquired plaintiff's information through improper means.

Plaintiff's information is a trade secret

Plaintiff's information is a trade secret. Restatement 1st of Torts s 757, and Comment b, factors 1–6 (1939). In the problem at hand, plaintiff was the only manufacturer making the product [F15]. This shows that the information apparently was not known or available outside plaintiff's business and plaintiff's information was valuable for plaintiff's business. In *Valco Cincinnati, Inc. v. N & D Machining Service, Inc.*, 24 Ohio St.3d 41, 492 N.E.2d 814 (1986), *Forest Laboratories, Inc. v. Formulations, Inc.*, 299 F.Supp. 202 (E.D.Wis.1969), *Mason v. Jack Daniel Distillery*, 518 So.2d 130 (Ala.Civ.App.1987), and *Digital Development Corporation v. International Memory Systems*, 185 USPQ 136 (S.D.Cal.1973), which held for plaintiff, there was similar evidence that plaintiff's information is a trade secret. In *Valco Cincinnati, Forest Laboratories, Mason*, and *Digital Development*, which held for plaintiff, plaintiff's product was different from products made by competitors [F15], as in the current case. The fact that plaintiff disclosed its information to defendant during negotiations [F1] does not necessarily rule out a conclusion that plaintiff's information is a trade secret. *Digital Development, Mason, Forest Laboratories, Valco Cincinnati.*

A confidential relationship existed between plaintiff and defendant

A confidential relationship existed between plaintiff and defendant. Restatement 1st of Torts s 757 (1939). In the current problem, defendant knew that plaintiff's information was confidential [F21]. This shows that defendant was on notice that using or disclosing the information would be a breach of confidentiality. In *Valco Cincinnati, Forest Laboratories, Mason*, and *Digital Development*, which held for plaintiff, there was similar evidence that a confidential relationship existed

Figure 1: Using CATO for making arguments about the *Steffi v. Vincent* hypothetical problem

3 Indexing Cases with SMILE

Despite these advantages, CATO has not made it into the law offices yet. One of the main obstacles for a wider use is the cost associated with adding new cases to the Case Database. The law is a continuously evolving domain, where new cases are decided every day. Further, decisions may be overturned on appeal, and these developments have to be incorporated into the casebase. It would be detrimental for an attorney if he missed a recent decision or used a case that was reversed on appeal. Similarly, for pedagogical reasons it is desirable to keep a tutoring system's casebase up-to-date. In our ongoing research, we try to address this problem and develop methods for automatically assigning Factors to new cases.

Our SMILE program takes an existing casebase as examples for how the indexing should be done in a machine learning (ML) approach. This research was motivated by the successful research in text classification in ML and IR; see (McCallum *et al.* 1998; Joachims 1998) for methods that establish a current performance standard. The approach in SMILE differs from these prevailing text learning methods. Rather than applying advanced statistical models suitable for very large sets of training examples, it uses a symbolic ML approach more suitable for our comparatively small collection of cases and integrates background knowledge for a better representation of the training examples (Brüninghaus & Ashley 1997; 1999).

An overview of SMILE is given in Figure 2. In the *Learning from Examples* phase, SMILE takes as input a set of annotated squibs, similar to the *Steffi v. Vincent* case. The marked-up sentences are positive training instances for a Factor, and all unmarked sentences are negative instances, whether they come from a case where the Factor applies or not. The sentences are represented as a set of features. SMILE uses a symbolic learning algorithm to induce rules or a classifier from which rules can be easily extracted.

The classifier can then be used for *Indexing of New Cases.* This includes classifying sentences from squibs, like those it

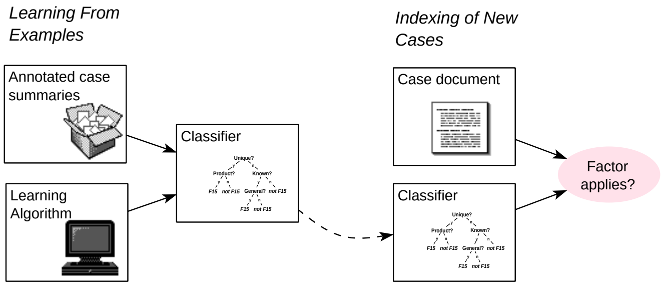


Figure 2: Overview of the learning approach and how new cases are indexed in SMILE

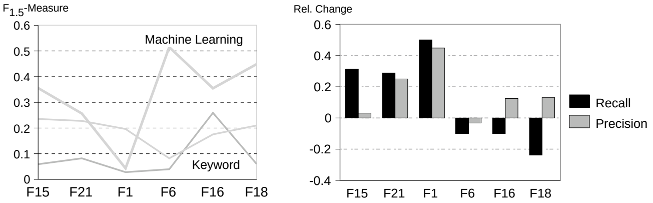


Figure 3: Results of the first evaluation of SMILE. The left graph compares the *Keyword* search for all or at least one word in CATO's Factor name to the *Machine Learning* approach, using the $F_{1.5}$ measure. The right graph shows the relative improvements of precision and recall when adding a thesaurus.

was trained on, and labeling sentences from the full-text opinions. The Factor is assigned to a new case text, if at least one sentence in the document was classified as a positive instance. Thus, SMILE can be used to bootstrap the development of large casebases with a small collection of annotated summaries.

In initial experiments reported in (Brüninghaus & Ashley 1999), we compared the way a human well-familiar with CATO would search for Factors to the classifiers learned with ID3. Humans often search for keywords from the descriptive Factor names. For instance, cases to be included in CATO's Case Database were retrieved with this strategy from WestLaw.

The results of the machine learning and keyword approaches were compared across Factors using the $F_{1.5}$ measure, which combines precision and recall in a single number, comparable to a weighted geometric average with a slight preference for recall. Figure 3 (left) shows that, apart from the particularly difficult Factor F1, the learning algorithm scored better than the keyword strategy often employed by humans.

The second part of the experiment tested the integration of background knowledge in the form of an online thesaurus from WestLaw to overcome the negative effects of overfitting. The implemented learning algorithm ID3 tends to learn overly specific trees; in particular it tries to cover all synonyms for one concept in individual branches, which can hurt performance. With the thesaurus, all variants referring to the same concept a covered by one decision tree node. Figure 3 (right) shows considerable performance improvements for some of the Factors (F15, F21, and F1), effects likely to offset each other for some Factors (F16 and F18), and for one Factor (F6) a minor performance decrease.

Summarizing, these experiments indicate that the ML approach is promising and that adding background knowledge can improve performance. However, the results are still far from perfect. A more detailed analysis of the data suggests that a more powerful representation of the examples is needed. The legal concepts underlying the Factors require a linguistic analysis of the examples and a better representation of knowledge about the language used in legal cases.

4 A Better Representation of Legal Text

Text classification research almost exclusively uses a bag-of-words representation. The text is split up into single words or word pairs; word order is ignored. When large amounts of training data are available, remarkable performance is accomplished by using advanced statistical learning algorithms. Recent research by West Group (Yang-Stephens *et al.* 1999) even shows that these methods can be beneficial in the legal domain, for indexing case abstracts under West's Key Number classification scheme.

However, for finding Factors, these statistical learning methods are not applicable, as demonstrated in our past experiments (Brüninghaus & Ashley 1997). In the West Group experiments, massive amounts of training data were available, as many as 20,000,000 already indexed abstracts. Moreover, assigning Factors to cases is a much harder problem than finding Key Numbers. The concepts underlying the Factors are significantly more specific and capture more of the contents of the individual case. We will illustrate below why single words do not suffice as features, and why a more powerful representation is needed.

In order to classify legal texts accurately, even though the concepts are more complex and training instances are scarce, we developed methods to find a more meaningful representation, which allows the learning algorithm to better generalize from the training examples. In the following section, we discuss how abstracting from the particular actors and events to their roles, capturing actions in more meaningful multi-word patterns and representing negation can help toward that goal.

4.1 Abstracting from Names to Roles

In legal cases, the parties, products, locations, etc., are usually referred to by their names. This is appropriate in court documents, but often prevents a classifier from generalizing from the examples, because names are specific to each case.[2] In the example sentence 'Steffi tells the method to her friend Vincent.', the names of both parties, Steffi and Vincent, would at best be ignored by a learning algorithm because they do not occur in any other cases. However, if this relevant information is discarded, the sentence becomes fairly worthless as an example of Factor F1. A better solution is to preserve this information, while at the same time making sure that the unique names are replaced by more general information. Abstracting from the specific actors and events of a case to their roles in the lawsuit will solve this problem. Examples of unique names can be found in CATO's squibs:

- Mason disclosed part of the recipe to Randle.
- Goldberg discussed and demonstrated his screw-in lead at Medtronic's headquarters.
- Hisel sent a letter to Chrysler explaining his idea for a glass-covered holder for license plates.

[2] Exceptions include very common names, like Smith, or large, research-oriented companies, like DuPont, which tend to be the target of industrial espionage.

With the original names, it is hard to find a general pattern in these examples. However, if we know more about the cases, we can substitute Mason, Goldberg, and Hisel with plaintiff; Randle, Medtronic, and Chrysler with Defendant; and recipe, screw-in lead, and idea for a glass-covered holder for license plates with information. Then, the sentences look as follows:

- Plaintiff disclosed the information to defendant.
- Plaintiff discussed and demonstrated the information at defendant's headquarters.
- Plaintiff sent a letter to defendant explaining the information.

Now, a very important pattern becomes obvious. In each of these sentences the plaintiff discloses information to the defendant. In fact, each of these sentences is an example for Factor F1 in CATO.

Apart from these intuitive benefits of abstracting from names, one can also make a more information-theoretic argument. In SMILE, single sentences are used as examples. Substituting names by generic roles brings the overall case context into these examples, thereby adding information content to the examples.

While trade secret opinions do not follow a standard format, the underlying fact situations follow a basic pattern and always contain certain elements. There is always a plaintiff, a defendant, and product-related information, the trade secret. This information belongs to the plaintiff, and it allegedly was learned in some way and used by the defendant.

In the first SMILE experiments, we also found empirical evidence why it is useful to substitute party names with their roles. In these experiments, overfitting was a major problem with the learning algorithm. The trees became overly sparse and unbalanced when ID3 picked (non-replaced) party or product names as attributes to distinguish positive and negative instances. This is an example of how individual names can even hurt performance.

4.2 Capturing Actions in Multi-word Features

For a human to decide whether a sentence should a be positive instance of a Factor, it is crucial to know who did something, or what was done by a particular person. Prevalent text learning methods do not support the linguistic analysis required to determine these facts and to capture them in more powerful, multi-features features. With a bag-of-words representation, there is no meaningful way to distinguish between the positive example of Factor F1, Disclosure-In-Negotiations, 'Plaintiff tells her information to her friend defendant' and the negative example 'Defendant told plaintiff he would treat her information as confidential.' The sentences would be represented as (defendant plaintiff tell information friend) and (defendant plaintiff information tell confidential treat), respectively. The only words not shared in this representation, 'friend', 'confidential', and 'treat', are not very decisive for the goal concept and should therefore not be considered by a classifier.

A better representation would capture the meaning of the Factor: Factor F1 applies if plaintiff disclosed the information. Since one can imagine various ways of disclosing or communicating something,[3] one good feature for the positive example would therefore be (π *disclose*),[4] where the italics indicate it is present whenever a synonym of disclose is found. It captures that plaintiff was the actor, and that she communicated or disclosed something. Further good features are (*disclose* τ) and (*disclose-to* δ). The negative example would have, among others, the feature (δ *disclose*). With this more meaningful representation, it will be easier to distinguish the examples and determine whether F1 applies.

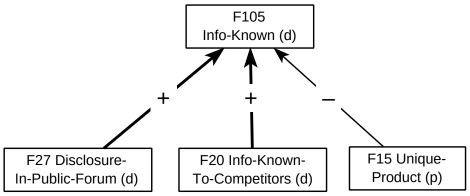

F105
Info-Known (d)
+
+
−
F27 Disclosure-In-Public-Forum (d)
F20 Info-Known-To-Competitors (d)
F15 Unique-Product (p)

Figure 4: Part of CATO's Factor Hierarchy, showing blocking relation between F15 and F20.

We call these features *propositional patterns* (ProPs). In order to derive the ProPs from examples, a linguistic analysis of the examples is necessary. In Section 5, we will discuss how a state-of-the-art NLP tool can be used to generate ProPs.

4.3 Dealing with Negation

Negation is another relevant feature for determining whether a legally relevant concept is present in a legal text. It has been shown (Riloff 1995) that small words, in particular 'no' and 'not', should not be removed as stopwords for text classification, because their presence can be relevant for certain concepts. These results are even more relevant for the law than for other domains, because legal discourse has an adversarial character. The court may consider evidence for and against a Factor in an opinion before presenting its conclusion. For instance, in the *National Rejectors* case, after discussing arguments presented by both parties whether the defendant knew that plaintiff's information was confidential, which is represented by Factor F21 in CATO, the court concludes 'There is *bf no* evidence that defendant knew that plaintiff had or claimed to have any trade secrets with respect to its marketed devices.' (emphasis added and names substituted)

Furthermore some Factors are defined to apply if certain facts are absent. Consider Factor, F15, Unique-Product, which applies when plaintiff's product is unique and not known in the industry. The definition of the Factor already indicates that the presence of negation can be decisive.

Negation becomes even more crucial if we also consider another Factor in CATO, F20, Info-Known-to-Competitors. This Factor represents a complementary situation to F15. When the product is not known to competitors, Factor F15 applies; when the information is generally known, Factor F20 applies.

This relationship between the factors ois represented in CATO's Factor Hierarchy. Figure 4 shows that Factors F15 and F20 are related to the higher-level issue F105, Info-Known. Factor F20 has a strong supporting link to F105, while F15 has a weak opposing link. This constellation is called *blocking*, where F20 blocks the conclusion supported by F15, namely that the information would not be known. While blocking is a consequence of the opposing meaning of Factors, Factors may also block each other even though they are not opposites. In Figure 4, Factor F27, Disclosure-In-

[3] (Rodale 1978) lists for instance reveal, tell, utter.

[4] In the remainder of this paper, we will use π, δ and τ to refer to the **p**laintiff, the **d**efendant and the **t**rade-secret-related information respectively.

Public-Forum blocks F15. Public disclosure generally implies that F15 does not apply. In order to correctly distinguish between examples of Factors F15 and F20, negation has to represented in the examples. This can be illustrated by the sentences 'Plaintiff's customers were well known to others in the trade.', which provides evidence for Factor F20 in *Springfield*, and 'The information in the customer list was not known.', which provides evidence for F15 in *Zoecon*.

We also found empirical evidence for the need to represent negation in our previous experiments with SMILE. For Factor F15, we analyzed for which subsets of training examples the learning algorithm could not find a good way to distinguish between positive and negative instances while it was constructing the decision tree. Quite often, the problem was that ID3 did not know about negation. We then added the words 'no' and 'not' to the representation and observed that the algorithm used these features in places where it had run previously out of useful features. While we did not formally evaluate these experiments, the observations suggest that adding negation is important.

4.4 Illustration: Better Classification with Improved Representation

To illustrate how a better representation could lead to better classification, consider CATO's Factor F1, Disclosure-In-Negotiations. It applies when plaintiff voluntarily disclosed the information to defendant, typically in business negotiations. In past experiments, F1 was particularly hard to find; overfitting occurred more often than for other Factors.

Figure 5 shows how an ideal (or at least close to ideal) classification tree would look. This particular tree was constructed by hand from the positive examples, including those presented in the example in Section 4.1.

The example representation in this tree includes the improvements suggested above. First, names of and references to plaintiff, defendant and the trade-secret-releated information were substituted by role identifiers π, δ, and τ, respectively. Then, the sentences were reduced to simple patterns, for instance 'π disclosed τ to δ.' from which the corresponding word pairs were derived. From this list, identifying four typical disclosure scenarios and constructing the classification tree in Figure 5 were straight-forward.

To show how this tree works for classifying sentences, consider a sentence from the *Steffi v. Vincent* example. 'Steffi tells the method to her friend Vincent.' is a positive example for the Factor F1, Disclosure-In-Negotiations. In this example, the specific names would be replaced by their roles, resulting in 'π tells τ to δ.' Assume we have a representation of semantics which contains that the communicative act of telling something is a disclosure of what one knows.[5] Next, we can use this generalization to derive the ProPs (π *disclose*), (*disclose* τ) and (*tell-to* δ) as features for the example. Because the subject is plaintiff, we focus on the leftmost branch. The example will pass the first test, because it has the features (π *disclose*) and (*disclose* τ). Therefore, the example will be correctly classified as an instance for Factor F1. The sentence 'Vincent tells Steffi that he will treat her method as confidential.' would be classified as negative. It has the subject defendant δ; we therefore focus on the third branch of Figure 5. Because this example has the feature (δ disclose), it will be handed down to the right through this branch and classified correctly as a negative instance for F1.

The resulting tree is compact and, as the example shows, fairly easy to understand. While it has not undergone a formative evaluation, this classifier can be expected to perform better than the decision trees learned in past experiments, which were much larger and harder to interpret. On unseen sentences, it will certainly accomplish high recall. Because the features are specific for Factor F1, one can expect fewer false positives, resulting in higher precision.

5 Implementation and Experiments

In order to learn the ideal classification tree discussed above, an implementation of the methods for abstracting from name to roles in the examples, for capturing multi-word features and for analyzing negation are required. While finding the parties or the product and abstracting from a sentence to word patterns is fairly easy for a human, automatically generating these features is a hard problem, because it requires background knowledge and a linguistic analysis of the documents.

5.1 AutoSlog

Before we continue to discuss how the better representation of examples can be implemented, we will first discuss the NLP tools used in SMILE. Up to now, NLP techniques have been considered inappropriate for legal case texts (see (Al-Kofahi, Grom, & Jackson 1999) for a notable exception), because the language used in legal documents is too complex. Sentences in the court's opinions are exceptionally long and often have a very complex structure. Consider for instance 'After a complete evidentiary hearing on plaintiff's claim for injunctive relief, the trial court concluded that Televation's schematics of its analog circuitry and the manner in which its analog circuitry interfaces with its digital circuitry are trade secrets.' (from the *Televation* case). Even for an English teacher, parsing this sentence will not be easy. While many problems are still far from being solved, recent progress in NLP has yielded tools that can handle some of the complexities of legal texts. We found that a state-of-the-art NLP package, the Sundance segmenter and the AutoSlog[6] information extraction (IE) program, developed by Ellen Riloff of the University of Utah, can handle most of the problems raised by legal cases. The following section focusses the most important functions of AutoSlog; a more in-depth discussion can be found in (Riloff 1996).

Given a sentence as input, the Sundance segmenter generates a (partial) parse. Sundance has very efficient heuristics to first break up a sentence into clauses, and then find the subject, verb, object and prepositional phrases. For an example of Sundance's output, see Figure 6.

Sundance is exceptionally robust and efficient. As with all parsers, its syntactic analysis of a sentence is often not absolutely correct. However, even where the parse contains a mistake, Sundance's output is usually still useful. In particular for the long sentences in legal cases, the segmenter tends to 'recover' on a later clause, which is not the case with many other parsers. A minor disadvantage of Sundance is the fact that the currently implemented heuristics were not designed for the rather long noun phrases in legal cases, which is the reason for most of the errors we encountered.

However, the most remarkable functionality of AutoSlog is its IE module. Given a target word, the program can find all linguistic contexts in which this word occurs in a document. To illustrate this, assume that the target word is 'method' and that the text is 'Plaintiff developed a method

[5] We applied this assumption in constructing the tree.

[6] Or, more accurately, AutoSlog-TS.

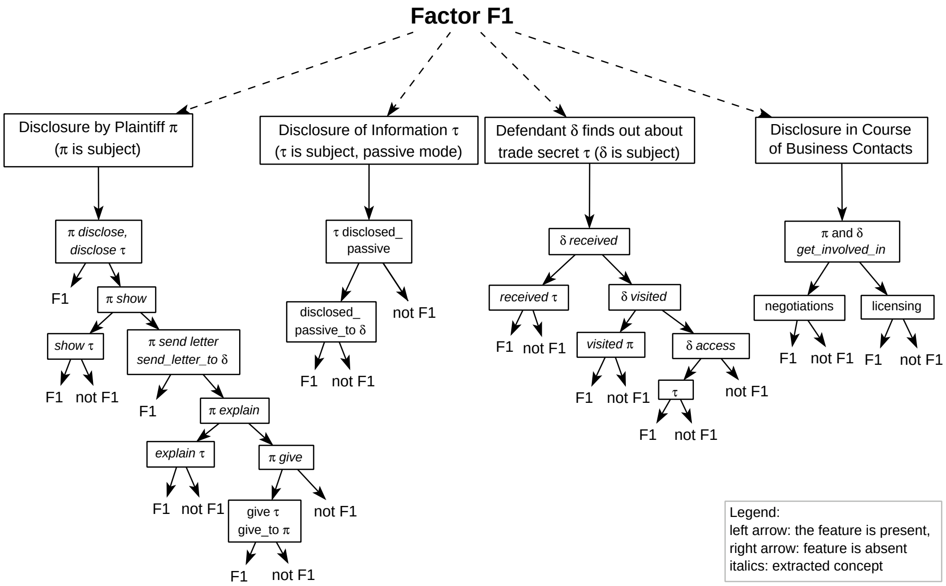


Figure 5: Manually derived classification tree for Factor F1, Disclosure-In-Negotiations

```
CLAUSE
  NP SEGMENT (SUBJ):
    [Plaintiff (LEX)(N SINGULAR(PLAINTIFF PERSON))]

  VP SEGMENT (ACTIVE_VERB):
    [developed (root: develop) (MOR)(V PAST(MANUFACTURE))]

  NP SEGMENT (DOBJ):
    [a (LEX)(ART)]
    [method (LEX)(N SINGULAR(PRODUCT))]

  PP SEGMENT (PREP):
    [for (LEX)(PREP)]
    NP SEGMENT:
      [indexing (INF-MOR)(GER)]
      [text (LEX)(N SINGULAR)]
      [cases (root: case) (MOR)(N PLURAL)]
  [>PERIOD (LEX)(PUNC)]
```

Figure 6: Sundance's output for the sentence 'Plaintiff developed a method for indexing text cases.'

for indexing text cases.' As we can easily see from the Sundance output in Figure 6, in the example sentence, 'method' is the direct object. The verb of the sentence is 'developed'. The context for the word 'method' in the example sentence is 'direct-object of the verb developed'. AutoSlog has heuristics to discover this context from the parser output of Sundance. These linguistic contexts are called *caseframes*. We will follow the notation in (Riloff 1996), and use developed_⟨dobj⟩ to refer to the caseframe 'direct-object of the verb developed'.

AutoSlog's caseframes can be used to extract information from new texts. To illustrate how AutoSlog's extraction mode works, consider the caseframe developed_⟨dobj⟩ ('direct-object of the word developed') and the sentence 'Plaintiff developed a method for indexing text cases.' As we know from Figure 6, the verb in the sentence is 'developed' and the direct object is 'method'.

In its extraction mode, AutoSlog first checks whether the caseframe is triggered by the input sentence. That is in our example, the program tests whether the verb in the caseframe is the same as the verb of the sentence. Because both have the verb 'developed', the caseframe is triggered by the sentence. In the next step, Autoslog extracts from the sentence the target indicated by ⟨⟩. In the example caseframe, the target is ⟨dobj⟩, the direct object. Therefore, the direct object 'method' would be extracted from the input sentence. The extracted information is also called the filler.

```
> Plaintiff developed a method for indexing text cases.
Caseframe DEVELOPED_01 (DEVELOPED; ACTIVE_VERB)
  DOBJ_EXTRACTION: "a method" [PRODUCT]: 1
```

Figure 7: From output of AutoSlog in extraction mode

Summarizing, given a caseframe and a sentence as input, AutoSlog extracts the target filler if the caseframe is triggered by the sentence. If the input sentence were 'Plaintiff developed after determined research a computer program', 'computer program' would be extracted as filler because the caseframe is triggered by the sentence.

Moreover, AutoSlog determines the semantic category of the filler. See Figure 7 for the output when the caseframe developed_⟨dobj⟩ is triggered by the sentence 'Plaintiff developed a method for indexing text cases.' After the word 'method' is extracted, AutoSlog infers from the entries in its trade secret law dictionary that the filler is a product for our application.

With this functionality, AutoSlog is a perfect tool for identifying and extracting information from CATO's case texts. The courts tend to repeat certain phrases. For instance, there are not too many ways to describe that a person, the plaintiff, developed a product. Thus, from a representative training set of products, like 'method', together with the sentences in which the product is referred to, like 'Plaintiff developed a method for indexing text cases.', one can derive a set of prototypical linguistic contexts, or caseframes, for the product, including developed_⟨dobj⟩. Other typical caseframes for the product are listed in Table 1. If one of those caseframes is found in a new case, it is reasonable to assume that the extracted filler information is a

Training Sentence	Product	Caseframe
SDRC began developing a computer program called NIESA.	computer program	developing ⟨dobj⟩
	NIESA	called:passive ⟨dobj⟩
The plaintiff manufactures adhesive tape, including a 'masking tape'	adhesive tape	manufactures ⟨dobj⟩
	masking tape	including ⟨dpbj⟩

Table 1: Examples of caseframes generated from squibs. The caseframes read: Extract ⟨syntactic-role⟩ if the sentence has the trigger verb.

product.

Of course, sometimes a caseframe extracts a filler, one was not looking for. Consider the sentence 'Plaintiff [aka Steffi] has a friend called defendant [aka Vincent].' Table 1 contains the caseframe called:passive_⟨dobj⟩, which is actually very effective for extracting product names. However, this caseframe would extract 'defendant', who is certainly not a product. The semantic analysis of the extracted filler from Figure 7 can help reduce these errors. It would have labeled 'defendant' as [PERSON]. Various statistical analyses over the training set can also help detecting overly general caseframes.

5.2 Finding the Parties in a Lawsuit

In Section 4.1, we showed three example sentences and how replacing the specific names of plaintiff and defendant leads to a more general representation that is more useful for finding Factors. Here, we present methods for implementing this technique and discuss how well they work.

As mentioned above, trade secret cases always have a party who owns the trade secret, and a party who allegedly misappropriated the secret. With respect to the trade secret claim (and for CATO), they are plaintiff and defendant, respectively. If the case is an appeal, the additional roles for the appeal may appear, but they will not affect the initial trade secrets claim.[7]

To find the parties from the full-text opinions, a number of clues from the cases can be used to infer the roles of the persons. If the opinion follows a standard format, the header will read 'Steffi, Plaintiff, v. Vincent, Defendant'. Half of the cases in CATO's Case Database have this pattern. Here, trivial pattern matching techniques suffice. Unfortunately, the other half of the cases requires more work. Various courts follow different stylistic guidelines. Similarly, when the case is an appeal, the parties are often not referred to as plaintiff and defendant in the header. In the majority of these cases, the court uses appositions to ascertain the parties roles somewhere in the body of the case; for instance, 'Plaintiff, Telerate Systems, Inc. ("Telerate"), ...', from the*Telerate* case. If it is possible to infer that 'Plaintiff' is the same as 'Telerate System, Inc.' and 'Telerate' in the parentheses, the names of the parties are known.

```
$list-of-names :     ($name, )*$name(,? and $name)?
$name :              $person-name | $company-name
$person-name :       $title? $first-name? $last-name | ...
$company-name :      $capitalized, $organization |
                     $person-name & Sons | ...
$first-name :        read from file
$last-name :         [A-Z][a-Z]* | [A-Z][a-Z]*-[A-Z][a-Z]*
$organization :      Inc. | Co. | Ltd. | ...
```

Figure 8: Sample grammar for names; $ indicates a variable, other meta-characters have the usual meaning (Friedl 1997).

Appositions are fairly difficult to handle correctly for an NLP program, because without additional clues, their attachment and scope may be ambiguous. For our task however, the names of the parties from the header of the case are known, we just do not know who is plaintiff and who is defendant, so we know the scope of the apposition. We also know that the head noun of the apposition is plaintiff or defendant, so we know the attachment. This makes the problem much easier to solve. We have identified and formalized three general patterns to cover appositions like in the *Telerate* example from the following sentence patterns:

- The plaintiff, $list-of-names, ...
- $list-of-names, plaintiff, ...
- Plaintiff $list-of-names ...

These patterns are appropriate for single names and lists of names. Our implementation can also recognize potential members of a party, which were only mentioned as *at al.* in the header of the case.

We used Perl (Wall, Christiansen, & Schwartz 1996) to implement the heuristics. This language offers particularly powerful and efficient regular expressions for dealing with strings and text (Friedl 1997). In Perl, patterns for various types of names and for appositions can be expressed easily in a rule-based grammar, which also facilitates matching name variations, for instance 'Mr. T. Smith' and 'Tom Smith'. Figure 8 is intended to give a flavor for the kind of rules and regular expressions we developed for names. These rules are not taken from the actual grammar, which is more complicated. We also use the grammar to guess whether a name refers to a person or a company.

In this first implementation, the heuristics work very well. Only for a small number cases, it is not possible to automatically determine the identity of the parties. In most of these cases, it is also somewhat difficult to determine the roles of the parties for a human. In particular where the court focuses on procedural issues, a deeper syntactic analysis of the cases may be required.

One method may be to use AutoSlog to improve upon the extraction of names with heuristic rules. In a trade secret claim, the plaintiff (by definition) sues, or brings suit against, the defendant. For instance, in the *Amoco* case, the header does not identify the plaintiff or the defendant, but we have a sentence 'Amoco [plaintiff] brought suit ... against Lindley [defendant].' Intuitively, a very indicative caseframe to determine the parties's roles is ⟨subj⟩_brought-suit (subject of a sentence that has the trigger 'brought' and the direct-object 'suit') which should extract the plaintiff with very high precision. When we tested this, all fillers (apart from a few parser errors caused by punctuation) extracted with this caseframe referred to the plaintiff. While most of the fillers were names, about 40% fillers are of the form 'Owner of computer software', which does not identify the name of the plaintiff. However, this filler can be used to

[7] We will not consider exceptions where the parties' roles are reversed, for instance a declaratory judgement or a trade secret counterclaim within a lawsuit. Because they are comparatively rare, the effort needed to cover them can not be justified in this project.

$$\left.\begin{array}{c}\text{Product-Names}\\ \text{Training-Squibs}\end{array}\right\} \xrightarrow{\text{AutoSlog}} \left.\begin{array}{c}\text{Product}-\\ \text{Caseframes}\\ \text{Test-Squibs}\end{array}\right\} \xrightarrow{\text{AutoSlog}} \begin{array}{c}\text{Extracted}-\\ \text{Products}\end{array}$$

Figure 9: Experiment to extract product names and product-related information

improve upon the methods for infering the product in this case, discussed below.

We are planning to set up a more in-depth evaluation. It is realistic to expect precision and recall to reach the 90 % range. This would compare to the results in Salomon (Moens 2000) for simpler structured cases.

5.3 Finding the Product

As outlined in the example in Section 4.1, another technique to improve the representation is to replace the name of products and product-related information by a more general term. In this section, we discuss why finding product-related information is hard, and how we use an IE program to overcome these problems.

Finding the alleged trade secret and product-related information is very difficult. First, there may be multiple aspects to the trade secret. For instance, the product sold in the market may be a car, while the trade secret is a unique machine to produce the car, and the information stolen by the defendant were the design drawings for that machine. For the purpose of finding Factors, all these aspects of the trade secret are relevant. Second, depending on the kind of secret, the way it is referred to can vary widely. The pattern matching methods used to find the parties would not be appropriate here. We would not be able to capture some of the different aspects of a trade secret. Because the wording in which the trade secret is referred can differ considerably, a more sophisticated syntactic analysis of the cases is necessary.

For extracting the product information, we plan to use AutoSlog. Given a set of target nouns, the program can generate the caseframes related to those nouns from input text. Examples of input sentences and prouct names with the corresponding caseframes are in Table 1. As mentioned above, caseframes can be interpreted as a form of extraction rule: if the trigger is present, extract the filler role. This extraction is not based on simple word matching, but on a linguistic analysis of the input case.

5.4 Evaluation and Experiments for Finding the Product

To determine whether we can use caseframes derived from a (training) set of examples, which comprises squibs with product names mentioned in these squibs, to extract product information from unseen (test) cases, we conducted an experiment, outlined in Figure 9. The following discussion will use the terms from Figure 9 for clarification. The experiment was carried out as 2-fold cross-validation (Cohen 1995).

We first manually extracted the trade secret-related noun phrases, or Product-Names, from every squib. For each training case, we gave the respective Product-Names and Training-Squib as input to AutoSlog. The program generated the Product-Caseframes for every training case. For examples of sentences from the Squibs, the manually extracted Product-Names and the Product-Caseframes, see Table 1. Most Product-Caseframes were very useful; however, parsing errors also lead to some unwanted instances. For instance, we got a caseframe which extracts the subject triggered by the verb 'signed', which will almost always return a person.

We then used the Product-Caseframes generated from the training set, as well as the Test-Squibs for the cases in the test set as input to AutoSlog in extraction mode. The output were the Extracted-Products. To catch some of the parser errors mentioned above, we filtered out all Extracted-Names that were automatically labeled by AutoSlog as a person, a contract or a date. For time reasons, we also removed extracted pronouns.

It should be pointed out that we did not delete any of the caseframes that were generated automatically from the training set. We used neither statistical thresholding nor did we manually go through the caseframes to make sure those generated by parsing errors would be removed. We only used information about the semantic category of the extracted noun phrase generated by the program where available.

From the cleaned-up list of Product-Names, we calculated recall as the percentage of products from manually extracted list of trade-secret related information that were found by the program. We did not count it as an mistake when part of the noun phrase was chopped off by the segmenter. Recall was defined as the percentage of correct references to the trade secret in the set of Product-Names extracted from a squib. We scored the extraction of generic terms, like machine or appliance, and references to defendant's products as correct. Even though manual scoring has always a certain bias, we tried to be as objective as possible in this evaluation.

In this experiments, precision and recall, macroaveraged over the cases, reached 66.15% ($\sigma = 0.12$) and 64.82% ($\sigma = 0.05$), respectively. Given that only half of the cases were in each training set split and that we performed no manual fine-tuning or statistical analysis to detect overly general or incorrect caseframes, we feel this a very good result. Finding product information is very difficult, and the language used to depends on the type of trade secret, whether it is a compilation of information or a marketed product. We are confident that further measures to filter out less useful caseframes will help increase precision.

Compared to the best reported performance at the MUC evaluations, the results for finding the product may not appear impressive. However, the participants of MUC sometimes spend man-years fine-tuning their systems to the particular extraction task, which would not be possible for SMILE.

5.5 Generating Propositional Patterns

As in the manual construction of the classification tree in Figure 5, the next measure to improve the example representation, generating ProPs, will take place after the names of and references to parties and products have been substituted by their roles in the lawsuit.

Recall the manually constructed tree in Figure 5. It has a node (π *show*), which would be present in sentences which have plaintiff as subject and a synonym of show as verb. This feature is very powerful because it corresponds to the goal concept: For Factor F1 to apply, the disclosure has to be made by the plaintiff or somebody acting on his behalf.

In a first step, we used AutoSlog to generate all possible caseframes for the sentences marked-up for F1 in the squibs. Table 2 lists the most frequent patterns from the squibs related to Factor F1. It also lists the ProPs corresponding to those caseframes, which can all be found in the manually generated decision tree in Figure 5. This table provides provides evidence how useful the caseframes are for generating the desired features.

Caseframe	Corresponding pattern
⟨subj⟩_gain	(δ gain)[8]
⟨subj⟩_give	(π give)
⟨subj⟩_show	(π show)
letter_to_⟨pp⟩	(letter_to δ)

Table 2: Most frequent caseframes from examples for F1, Disclosure-In-Negotiations, and the corresponding Propositional Patterns.

The next part of the implementation will generate the ProPs from the caseframes. For this, we will first extract the filler noun phrases from the example sentences. In the next step, we will generalize from these fillers to more general concepts, using the semantic analysis performed by AutoSlog. In Figure 7 for instance, the semantic label [PRODUCT] for the filler 'method' was inferred automatically by the program. We are in the process of representing the relevant language in a semantic hierarchy, so that it can be used to determine the best semantic label for the filler by AutoSlog.

5.6 Determining the Scope of Negation

The last technique to improve the representation of examples for finding Factors is the detection and representation of negation.

Negation is one of the harder problems in NLP. The presence or absence of the words 'no' or 'not' in a sentence is not equivalent with the presence or absence of negation. Furthermore, even if 'no' or 'not' are spotted, this does not indicate what was negated. To illustrate why the scope of negation is important, consider a sentence from the *Steffi v. Vincent* example, 'Because they were friends, the defendant did not sign a non-disclosure agreement.' For deciding that this sentence should not be classified under Factor F4, Non-Disclosure-Agreement, it is important to know what is negated. Only a linguistic analysis of the sentence can determine the scope of the negation. Consider a variation of the example, where we move the 'not' five words to the left: 'Because they were not friends, the defendant did sign a non-disclosure agreement.' This sentence has to be classified under F4. Unless we can not only ascertain that there is negation, but what is negated, it impossible to correctly classify the two examples.

Our solution is to use Sundance to determine the likely scope of the negation in the examples. The Sundance segmenter reliably splits a sentence into clauses before analyzing it syntactically. The example would be split into the clauses 'they were friends' and 'the defendant did not sign a non-disclosure agreement'. These clauses are the typical scope of negation in the squibs. Therefore, we can assume that all features, in particular the ProPs, derived from a clause which contains a negation are covered by that negation.

6 Summary

In this paper, we presented our research on automatically indexing legal case texts with ML methods. We argued that existing methods, both for extracting information from legal texts and for document classification are not powerful enough for finding indexing concepts in cases. From the analysis of our past experiments, we identified three hypotheses how a more powerful representation of the text cases will lead to improved indexing.

A learning algorithm will better generalize from examples and classify new cases if the representation (1) abstracts from the specific names of the parties and products to their roles in the lawsuit, (2) captures events and actions in multi-word features, and (3) recognizes negation. In a hand simulation, we showed that these methods lead to a more powerful classifier. We presented the design for an implementation, and evaluated the first modules. The experiments indicated that domain-specific heuristics are appropriate for finding the parties and that a state-of-the-art IE tool can be applied to detect the product.

In the coming months, we will continue to implement the outlined methods for deriving multi-word features and representing negation. In particular, we have planned a comprehensive evaluation of the representation with various experiments on different sets of cases.

If successful, the methods presented here could be beneficial for other applications, beyond case indexing. It has long been recognized that a better text representation could lead to significantly improved information retrieval systems (Turtle 1995), yet all approaches that integrated NLP tools to derive phrases or used thesauri to find synonyms did not lead to the expected, consistent performance improvements; see for instance (Voorhees 1998). The improved representation for SMILE is more likely to lead to improvements than the domain-independent methods tried in the past, which were often too general. The methods discussed here were developed specifically for legal cases and for the kind of concepts relevant in the legal domain.

If the cases in a full-text retrieval system were represented as suggested here, more abstract and powerful queries and quasi-conceptual retrieval would be possible. For instance, instead of submitting a query for 'disclos*' and 'plaintiff', which would return almost every trade secret case, he could ask for cases where plaintiff did disclose something, that is, for cases which have the ProP (π disclose).

Acknowledgements

This research has been supported by the National Science Foundation under Grants IRI96-19713 and IRI99-87869. We are very grateful to Ellen Riloff from the University of Utah for letting us use her AutoSlog program. We also thank Vincent Aleven from Carnegie Mellon University for his comments on this paper and for being a friend who would never missapropriate a trade secret.

References

Al-Kofahi, K.; Grom, B.; and Jackson, P. 1999. Anaphora resolution in the extraction of treatment history language from court opinions by partial parsing. In *Proceedings of the Seventh International Conference on Artificial Intelligence and Law*, 138–146.

Aleven, V. 1997. *Teaching Case-Based Argumentation through a Model and Examples*. Ph.D. Dissertation, University of Pittsburgh.

Ashley, K., and Aleven, V. 1997. Reasoning Smbolically about Partially Matched Cases. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*, 335–341.

Ashley, K. 1990. *Modeling Legal Argument, Reasoning with Cases and Hypotheticals*. MIT-Press.

Branting, L. 1999. *Reasoning with Rules and Precedents - A Computational Model of Legal Analysis*. Kluwer Academic Publishers.

Brüninghaus, S., and Ashley, K. 1997. Using Machine Learning for Assigning Indices to Textual Cases. In *Proceedings of the Second International Conference on Case-Based Reasoning*, 303–314.

Brüninghaus, S., and Ashley, K. 1999. Toward Adding Knowledge to Learning Algorithms for Indexing Legal Cases. In *Proceedings of the Seventh International Conference on Artificial Intelligence and Law*, 9–17.

Cohen, P. 1995. *Empirical Methods for Artificial Intelligence*. MIT Press.

Daniels, J., and Rissland, E. 1997. Finding Legally Relevant Passages in Case Opinions. In *Proceedings of the Sixth International Conference on AI and Law*, 39–46.

Friedl, J. 1997. *Mastering Regular Expressions*. O'Reilly & Associates, Inc.

Joachims, T. 1998. Text Categorization with Support Vector Machines: Learning with many relevant Features. In *Proceedings of the European Conference on Machine Learning (ECML-98)*.

McCallum, A.; Rosenfeld, R.; Mitchell, T.; and Ng, A. 1998. Improving Text Classification by Shrinkage in a Hierarchy of Classes. In *Proceedings of the Fourteenth International Conference on Machine Learning*, 359–367.

Moens, M.-F. 2000. *Automatic Indexing and Abstracting of Document Texts*. Kluwer Academic Publishers.

Prakken, H., and Sartor, G. 1998. Modelling Reasoning with Precedents in a Formal Dialogue Game. *Artificial Intelligence and Law* 6:231–287.

Riloff, E. 1995. Little Words Can Make a Big Difference for Text Classification. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

Riloff, E. 1996. Automatically Generating Extraction Patterns from Untagged Text. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, 1044–1049.

Rissland, E.; Skalak, D.; and Friedman, T. 1996. BankXX: Supporting Legal Arguments through Heuristic Retrieval. *Artificial Intelligence Review* 10:1–71.

Rodale, J. 1978. *The Synonym Finder*. Warner Books.

Turtle, H. 1995. Text Retrieval in the Legal Word. *Artificial Intelligence and Law* 2(1):5–54.

Voorhees, E. 1998. Using WordNet for Text Retrieval. In *WordNet: An Electronic Lexical Database*. MIT Press. 285–303.

Wall, L.; Christiansen, T.; and Schwartz, R. 1996. *Programming Perl, Second Edition*. O'Reilly & Associates, Inc.

Weber, R. 1999. Intelligent Jurisprudence Research: a new concept. In *Proceedings of the Seventh International Conference on AI and Law*, 164–172.

Yang-Stephens, B.; Swope, M.; Locke, J.; and Moulinier, I. 1999. Computer-Assisted Classification of Legal Abstracts. In *Proceedings of the Third International Symposium on Advances in Intelligent Data Analysis*, 437–448.