MEMO
To: Students in AI and Legal Reasoning Seminar
From: Kevin Ashley
re: Collaborative Design Problem
Date: January 3, 2017

Students,

This is a draft that I prepared recently in connection with a proposal for sabbatical funding for spring 2018. It describes the kind of project that I would like us to focus on for the seminar Collaborative Design Problem.

Please read this for the first class on January 11, 2017 and we can discuss it.

Best,
KA
ANNOTATING LEGAL CASES AND STATUTES FOR ARGUMENT MINING AND PEDAGOGY

New text analytic techniques promise to revolutionize legal information retrieval, but the machine learning on which they depend requires sets of training instances comprising manually annotated legal cases and statutes. "Annotating" means marking-up the texts of case decisions or statutes to identify instances of semantic types of information that are both important for conceptual legal information retrieval and pedagogically relevant. In order to support this need for manual annotation of legal texts, it is possible that students learning law could perform these annotations as part of their studies. As a by-product, they would produce the semantically annotated legal texts with which machine learning programs could be taught to automatically perform similar annotation of new texts.

AN OPPORTUNITY AND AN OBSTACLE

Recent research and developments in question answering, information extraction, and argument mining from text have planted the seeds of this revolution with programs like IBM’s Watson and Debater and their underlying open-source information management architectures. These programs will not perform legal reasoning themselves, but their open source text analysis tools will enable the development of new legal applications. The machine learning based tools will identify argument-related information in legal texts that can transform legal information retrieval into a new kind of conceptual information retrieval: argument retrieval. As a result, lay and professional legal practitioners will be able to retrieve information more effectively from legal databases. Conceptual legal information retrieval addresses the need in common law, and increasingly in civil law, jurisdictions to retrieve legal case opinions whose relevance lies in how they can be used in particular argument schemes for analyzing new problems.

While new legal apps such as Lex Machina,\(^1\) Ross,\(^2\) and Ravel\(^3\) employ text analytics and some can even make predictions, they do not address the substantive merits of a case and cannot make legal arguments or explain their predictions because they lack computational models of legal reasoning or argument. The computer science research field of Artificial Intelligence and Law (AI & Law) provides such computational models, and semantic annotation helps to fill the gap between legal case texts and those models of legal reasoning, enabling a computer to recognize, understand and process elements of a court’s reasoning.

The newly extracted argument-related information enables the computational models of legal reasoning and argument to deal directly with legal texts. These models will perform legal reasoning; they can generate arguments for and against particular outcomes in problems input as texts, predict a problem’s outcome, and explain their predictions with reasons that legal professionals will recognize and can evaluate for themselves. The result will be a new kind of legal app, one that enables cognitive computing, a kind of collaborative activity between human and computers in which each does the kind of intelligent activity that it can do best. For particular domains, it will support cognitive computing legal applications that can help humans to pose and test legal hypotheses, predict outcomes, and explain the predictions.

For argument mining to succeed, however, the machine learning programs need sets of training instances manually annotated by humans. The legal decisions can be annotated in terms of certain roles that sentences play in legal argument. The sentence annotation types

---


include stating a legal rule, expressing a judge’s holding that a rule requirement has been satisfied (or not), reporting a finding of fact, describing evidence, and others. In addition, the decisions can be marked-up in terms of general structural features of arguments such as premises and conclusions as well as substantive features of legal domains. Such features include legal factors, patterns of fact that strengthen or weaken a side’s position on a claim.

In related work, researchers are developing annotation schemes with pedagogical potential for marking-up statutes.

A substantial obstacle, however, involves the question of who will perform the manual annotation of training sets. While crowd sourced text annotation may succeed with some kinds of texts, it is likely that marking up legal cases and statutes requires annotators with some level of legal expertise. Legal professionals, however, are often unwilling to devote substantial amounts of time to the task.

**SOLUTION VISION**

From an educational viewpoint, I hypothesize that students of law could learn valuable lessons from annotating legal texts. The annotating task would draw students’ attention to key aspects of the reasoning in a legal case: the roles that certain sentences play in legal argument, structural features of argumentation, and substantive strengths and weaknesses of a legal argument involving a particular area of law. Indeed, in US law schools, much of the first year curriculum is aimed at inculcating skills of identifying these aspects as students learn to “read a case.” Annotation exercises supported by new technological techniques could fit well into the existing law school curriculum and reinforce these important lessons. As they learn, the students would produce useful data with which machine learning programs can learn to annotate texts automatically.

Three components are necessary in order to support students in annotation as a learning activity.

First, in order to incorporate annotation into the curriculum for law students (graduate students, undergraduates and even high school students) one needs to develop convenient web-based mark-up environments that could be used on tablet computers or laptops and that would make annotation almost as convenient as highlighting texts on line.

Second, the annotation environments must be part of an online system that ensures annotator reliability (i.e., agreement) across multiple students annotating the same documents. Maintaining reliability is important for enabling successful machine learning.

Third, one needs to tailor the annotation activities into a legal curriculum by developing a set of pedagogical materials that not only guide students in annotation but also underscore the lessons to be drawn from the annotation experience. This may include computer-supported peer review, in which student annotators critique each other’s annotations, as well as formative evaluation instruments for assessing whether and what students have learned from the annotation.

**NEED FOR DAAD SUPPORT**

My goals during the sabbatical period are to strengthen my connections to and research collaborations with particular German research groups whose technology and insights will be instrumental in realizing the vision, especially regarding the first two of the above components. DAAD’s support will enable me to pursue these collaborations. In particular, I hope to spend time in extended discussions with the following groups:

- Ubiquitous Knowledge Processing (UKP) Lab, Dept. of Computer Science, Technische Universität Darmstadt (TUDA), Dr. Iryna Gurevych, Director.

---

• Institute of Philosophy, Karlsruhe Institute of Technology (KIT), Prof. Gregor Betz
• Software Engineering for Business Information Systems, Department of Informatics, Technische Universität München, Profs. Dr. Florian Matthe and Bernhard Wald.

The UKP Lab in Darmstadt has developed the WebAnno tool, a convenient on-line annotation environment that supports annotation through crowdsourcing. My research colleagues are already using WebAnno in small-scale annotation applications and we would like to extend it into a pedagogical environment.

The KIT group is working on development of text corpora for policy argumentation and of argumentation schemes that extend surface annotations into deeper interpretative argument analyses.

The TUM team is focused on annotating more indirect, inferential citation references in legal texts and on semantic analysis of legal statutory and regulatory texts in connection with business compliance. Since annotation of legal documents (e.g., contracts, briefs, and memoranda) could also be an important training activity for in-house counsel, legal associates and paralegals, the Munich team’s experience applying annotation in corporate legal settings could lead to mutually beneficial collaborations.

BACKGROUND AND RELATED WORK

My book, Artificial Intelligence and Text Analytics: New Tools for Law Practice in the Digital Age, to be published by Cambridge University Press in late spring 2017, explains how the text analytic technologies enable new tools for legal practice using computational models of legal reasoning and argumentation developed by AI & Law researchers. It introduces the techniques for automated question answering, information extraction, and argument mining from texts as well as various AI & Law computational models of legal reasoning and argument that the text analytic technologies will connect directly to legal texts.

The computational models illustrate how to represent legal cases so that a computer program can reason about whether they are analogous to a case to be decided. They illustrate ways in which a program can compare a problem and cases, select the most relevant cases, and generate legal arguments by analogy for and against a conclusion in a new case. Some of the computational models of legal analogy integrate values and policies into the measures of case relevance, predict case outcomes, and explain the predictions in terms of arguments.

The Artificial Intelligence and Text Analytics book introduces a specialized kind of ontology for information extraction, “type systems,” which are a basic text analytic tool. Type systems support automatically marking-up or annotating legal texts semantically in terms of important concepts. It explains how case and statutory texts are represented for purposes of applying machine learning, natural language processing, and manually

---

6 See, e.g.,

7 Notably, Grabmair’s VJAP and Atkinson’s and Bench-Capon’s VAP models, ibid.
constructed rules to extract information.\textsuperscript{8}

<table>
<thead>
<tr>
<th>Citation</th>
<th>Sentence includes a citation to a legal authority.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legal rule</td>
<td>Sentence states a legal rule in the abstract, without applying it to particular facts.</td>
</tr>
<tr>
<td>Legal rule requirement</td>
<td>Sentence that states a requirement or element of a legal rule in the context of applying the requirement to the facts of the particular case being litigated.</td>
</tr>
<tr>
<td>Legal ruling or holding of law</td>
<td>Sentence states a legal ruling or holding of law by the judge.</td>
</tr>
<tr>
<td>Evidence-based finding of fact</td>
<td>Sentence reports the fact finder’s finding on whether or not evidence in a particular case proves that a rule condition has been satisfied.</td>
</tr>
<tr>
<td>Evidence</td>
<td>Sentence summarizes an item of evidence in the case.</td>
</tr>
<tr>
<td>Evidence-based intermediate reasoning</td>
<td>Sentence involves reasoning about whether evidence in a particular case proves that a rule condition has been satisfied.</td>
</tr>
<tr>
<td>Legal policy</td>
<td>Sentence states a legal policy or value in the abstract without applying it to particular facts.</td>
</tr>
<tr>
<td>Applied legal value</td>
<td>Sentence that involves reasoning about the application of a legal policy or value to particular facts.</td>
</tr>
<tr>
<td>Legal factor</td>
<td>Sentence in which a judge states as a reason why, or despite which, s/he came to a conclusion of law that a legal rule or requirement did [not] apply in a fact situation and refers to a stereotypical fact pattern that tends to strengthen or weaken the legal conclusion because of its effect on a legal policy or value.</td>
</tr>
<tr>
<td>Policy-based reasoning</td>
<td>Sentence involves reasoning about the application of a legal policy to particular facts.</td>
</tr>
<tr>
<td>Case-specific process or procedural facts</td>
<td>Sentence refers to the procedural setting of or a procedural issue in the case.</td>
</tr>
</tbody>
</table>

Figure 1: LUIAMA sentence level types that law students need to understand in learning to read a case.

These techniques have automatically identified argument-related information about some of the roles of sentences as, for example, statements of legal rules in the abstract or as applied to specific facts, or as case holdings and findings of fact.\textsuperscript{9} Figure 1 shows an extended list of the LUIAMA sentence level types. Law students would benefit from practice identifying the sentences playing these roles in legal cases; their annotations of such roles could enable machine learning programs to learn to successfully extract the remaining roles.

Information extraction techniques also have automatically annotated more general roles such as propositions in arguments, premises or conclusions,\textsuperscript{10} and the argument schemes that justify the conclusions given the premises, schemes such as analogizing the current facts to a prior case or distinguishing them.\textsuperscript{11}

\textsuperscript{8} Ashley, op. cit.


Finally, some progress has been made in machine learning to identify information that affects the strength of an argument such as evidence factors, a fact-finder’s stated reasons and assigned plausibility-values for a conclusion, or legal factors, stereotypical fact patterns that strengthen a particular type of legal claim. For example, Figure 2 refers to three such legal factors, F6 Security-Measures, F15 Unique-Product, and F19 No-Security-Measures, that apply in the domain of claims for trade secret misappropriation, as well as their effects on certain policy values that trade secret law protects.

The LUIMA architecture supports applying a type system and text annotation pipeline to process case texts for argument-related information about sentence roles. It employs the techniques for automating conceptual markup of documents and for extracting information from legal case texts and integrates them into a prototype system for conceptual legal information retrieval. The system has modules for automatic sub-sentence level annotation, machine-learning-based sentence annotation, basic retrieval using a full-text information retrieval system, and a machine-learning-based re-ranking of the retrieved documents. The automated annotation and machine learning are based on manually annotated training sets of documents. Evaluations have demonstrated objectively the contribution the system’s argument-related information makes to improve the full-text legal information system’s rankings.

Thus, evidence supports the LUIMA hypothesis, that argument retrieval is feasible. By semantically annotating documents with argument role information and retrieving them based on the annotations, one can outperform current systems that rely on text matching and current techniques for legal information retrieval. The Artificial Intelligence and Legal Analytics book explores how to extend the LUIMA type system to other kinds of conceptual queries and to statutory information retrieval.

Given robust argument retrieval, an app could support a cognitive computing environment tailored to the legal domain in terms of tasks, interface, inputs and outputs. Type systems and annotations based on the computational models could help humans frame hypotheses about legal arguments, make predictions, and test them against the

Figure 2: Some policy values underlying trade secret law (italics) and selected factors affecting them (bold).

- Plaintiff’s property interest is legitimate
  - because plaintiff’s information was unique in that plaintiff was the only manufacturer making the product. (F15 Unique-Product)
- Plaintiff has protected his property interest
  - because plaintiff took active measures to limit access to and distribution of its information. (F6 Security-Measures)
- Plaintiff has protected his confidentiality interest
  - because plaintiff took active measures to limit access to and distribution of its information. (F6 Security-Measures)
- Plaintiff has waived his interest in confidentiality
  - because plaintiff failed to take active measures to limit access to and distribution of its information. (F19 No-Security-Measures)


14 Grabmair, et al., op cit; Bansal, et al. op. cit.
15 Ibid.
16 Ashley, op. cit., Chapters 11, 12.
documents in a corpus. Posing and testing legal hypotheses is a paradigmatic cognitive computing activity in which humans and computers can collaborate, each performing the kind of intelligent activity it performs best. Humans know the hypotheses that matter legally; the computer helps them to frame and test these hypotheses based on arguments citing cases and counterexamples. The type system annotations will enable a conceptual legal information system to retrieve case examples relevant to the hypotheses, generate summaries tailored to the users’ needs, construct arguments, and explain predictions.17

**SOLUTION OVERVIEW**

By the end of the spring term in 2017, I hope to have gained some insight and experience by engaging law and graduate students in designing and applying an extended pedagogical activity of annotating legal texts. The *Artificial Intelligence and Legal Analytics* book will serve as the course textbook for my spring term seminar on AI & Legal Reasoning at the University of Pittsburgh School of Law. As part of the seminar, students will engage in designing and trying out exercises that focus students on annotating sentence roles and legal factors in trade secret misappropriation cases. Later in the course, students will observe how their annotations work in conjunction with the LUIMA architecture to enable the Value Judgment-based Argumentative Prediction (VJAP) computational model to make arguments about and predictions for the cases they have annotated.

![Figure 3: WebAnno annotation environment for marking up trade secret factors.](image)

**ANNOTATION ENVIRONMENT:** In particular, this will be our first attempt at using UKP’s WebAnno to orchestrate a larger scale annotation effort as a pedagogical activity, employ

---

17 Ashley, *op. cit.*, Chapter 12.
multiple annotators per case and monitor reliability of annotation. As suggested in Figure 3, students read a case on screen and highlight sentences they label as instances of particular factors. A manual will guide them through the process of annotating factors, including providing a list, definitions, and examples of the 26 trade secret factors that have been identified so far as well as sentence argument roles such as stating findings of fact and legal conclusions.

We will explore: (1) Motivating student annotators to achieve higher productivity and reliability of annotation by introducing some competition. (2) Employing computer-supported peer review through which students can critically review each other’s annotations in order to achieve reliability.

**DEMONSTRATION OF REASONING**: Once students have annotated some cases, we will demo inputting them into the VJAP program, which can predict an outcome and generate arguments explaining the prediction. VJAP employs a legal domain model of trade secret misappropriation law shown in Figure 4. The upper part comprises the rules defining a trade secret claim, including the rules’ terms or issues, such as the requirement to Maintain Secrecy. Each issue has associated factors that can strengthen or weaken a side’s argument that the issue has been satisfied.

In realistic legal disputes, it is frequently true that some facts favor a side and others favor an opponent. In Figure 3, for instance, the student has marked instances of the three factors illustrated in Figure 2: F6 Security-Measures, F15 Unique-Product, and F19 No-Security-Measures. F6 and F15 favor the plaintiff, while F19 favors the defendant. In another case called Dynamics, the factors also conflicted; F15, F6 and F4 (Agreed-Not-To-Disclose) favored the plaintiff, while F27 Disclosure-In-Public-Forum and F5 Agreement-Not-Specific favored the defendant.

Based on its domain model, Figure 4, the VJAP program determines the issues in the rules defining a trade secret claim that are affected by the conflicting factors. It employs additional information based on Factors’ effects on values protected under trade secret regulation as in Figure 2. In light of these effects on values, deciding an issue or the claim for one side or the other protects certain interests at the expense of other interests.

The VJAP program resolves tradeoffs into confidence values in an argument graph and aggregates them quantitatively using the domain model. The arguments in the graph are constructed with schemes for making and responding to arguments by analogy that a legal rule should apply to a current case based on past cases. These analogies assert that the current case and prior cases present the same (local or inter-issue) tradeoffs in value effects and should therefore have the same results. Using these schemes the program can retrieve...
cases sharing the same tradeoffs as a current case and apply them in textual arguments. A portion of its argument for the Dynamics case is shown in Figure 5. The underlined phrases in the text illustrate where the argument scheme introduces tradeoffs in effects on values protected by trade secret law. Here, VJAP draws an analogy to the National-Rejectors case, which involves similar tradeoffs in value effects.

The argument graph represents all possible arguments about who should win the case given VJAP’s domain knowledge, argument schemes, and the other cases in its corpus of 121 trade secret cases. Figure 6 illustrates an excerpt of the argument graph it generates for the Dynamics case. A quantitative model associated with the argument graph predicts the outcome in the new case given the value tradeoffs in the previous cases.

The model propagates quantitative weights across the graph. The weights correspond to the degree of confidence with which the argument premises can be established and that the prediction is correct. These depend on the strength of arguments pro and con the premises, which, in turn, depend on the magnitude of promotion or demotion of the value in past case contexts. The confidence measure is increased in relation to the strength of the analogy between a precedent and the current case and decreased to the extent they can be distinguished.

In an evaluation of VJAP, the system learns in a training step the optimal fact effect weight parameters to maximize prediction accuracy. It uses simulated annealing, a technique for finding the global maximum of a function like confidence while avoiding local maxima. VJAP achieved an accuracy of 79.3% versus a majority label baseline of 61%.

PEDAGOGICAL IMPACT: From an instructional viewpoint, students and I will brainstorm designing the...
pedagogical annotation environment. They will receive hands-on experience with annotation tasks and an opportunity to see how these enable a computer program to make predictions and arguments. The *Artificial Intelligence and Legal Analytics* book provides an overview and selected details on each of the steps in this process. Students will also learn about the relevant metrics and approaches for evaluation in connection with each step. I will prepare pre- and post-test formative instruments to assess the pedagogical effectiveness in terms of learning relevant concepts of argumentation and trade secret law. Between now and my sabbatical, I will be working to prepare the mechanisms and infrastructure for such pedagogical annotation efforts.

**PROPOSED DAAD WORK:**

With DAAD support, I will spend one or two weeks visiting each of the above groups in Germany. Based on the experience of the student design and annotation activity in spring 2017, I will present seminars and demonstrations. I will provide useful feedback to the UKP Lab concerning applying WebAnno to organize a course-focused instructional annotation activity, engage in discussions with the KIT group about corpus development for interpretive arguments concerning policy- and valued-based reasoning, and learn from the TUM team how the techniques could be adapted to a more business-oriented annotation exercise involving legal professionals. In particular, I expect that the discussions will focus on the following.

**UKP: ADAPTING ANNOTATION ENVIRONMENTS TO LEGAL TEXTS**

From the viewpoint of annotation and modeling, legal texts evidence an explicit argumentation structure and a predictable set of semantic types, such as the sentence roles in Figure 1. This is one reason why it makes pedagogical sense to teach law students about these structures by giving them practice in annotating them. These structures include appellate argument and reasoning about evidence, for example, about causation. Other domains like scientific argumentation evidence similar structures, for example, sentences stating a hypothesis or distinguishing previous studies from the current work.

Such structures and types are more domain-specific than the proposition/conclusion structures on which previous argument mining work has focused. The more general structures may be instrumental in helping to identify domain-specific structures and types, but the latter will support the actual domain-focused conceptual information retrieval. For example, sentences that state trade secret factors are likely to be propositions in a legal argument whose conclusion is a legal claim, and more likely to be contained in a court’s statement of a finding of fact. Those more general argument types may increase a program’s confidence in assigning a factor to a new case, but it is the assigned factor that is more useful in helping users find a more relevant decision text.

A general topic for discussion with the UKP team will be the extent to which WebAnno supports annotation of instances of these more domain-specific structures and semantic types. The experience of the seminar students will indicate if the current level of support is adequate or if and how it might be improved. For instance, the multiple semantic types increase the degree to which annotations overlap, which can lead to a confusing degree of visual clutter.

In addition, given some of the semantic types in Figure 1, such as evidence-based intermediate reasoning, legal policy, applied legal value, policy-based reasoning and, to some extent, legal factor, are likely to be more subjective and to challenge reliability of annotation. In this respect, they are similar to the subjectively defined labeling criteria in

---


the IBM Debater project. Discussion with the UKP team may focus on the level of support WebAnno provides for flagging and resolving specific disagreements about such annotations.

In this connection, computer-supported peer review may play a role. For example, the web-based SWoRD system (Scaffolded Writing and Rewriting in the Disciplines, http://sword.lrcd.pitt.edu) implements reciprocal peer review of writing. Instructors’ detailed criteria guide students in critiquing each other’s essays. SWoRD scaffolds a cycle of writing, reviews, back-reviews, and rewriting and performs a variety of statistical reliability analyses. I would like to explore with the UKP team the extent to which peer-review could be adapted to the task of helping students to critique each other’s annotations, identify disagreements, and resolve them, improving reliability in the process.

KIT: INTERPRETATION AND THE LIMITS OF ANNOTATION

I expect that Dr. Gregor Betz and I will discuss the potential for my pedagogical approach to annotation, and its limitations, in leading students to go beyond surface-level annotations and engage in deeper interpretative argument. Interestingly, some of the sentence role annotation types in Figure 1 relate to legal policies and values and their application to specific factual scenarios. These annotations indicate where content relevant to interpretive argument is located, although they do not necessarily provide a human or a program with much information about the meaning of that content. The claim-specific legal factors, however, and their associations with effects on underlying values, Figure 2, do provide such semantic information. Thus, the surface level annotations to some extent connect to elements of the deeper interpretive reasoning on which the KIT team focuses. The VJAP program provides an example of that extension into deeper interpretation. Of course, VJAP takes advantage of certain constraints in legal reasoning and argument that are not generally applicable to the social and scientific policy debates in which the KIT team is most interested. Nevertheless, I expect there will be much to discuss.

TUM: STATUTORY AND REGULATORY ANNOTATION FOR BUSINESS COMPLIANCE

The seminar annotation activity will focus primarily on annotating case decisions, where we have elaborated a fairly comprehensive type system as in Figure 1. Annotating statutes and regulations is another application that has important ramifications. There are practical ramifications for business compliance and pedagogical ones, too, since American law schools often do not adequately prepare students to understand and analyze statutes. Discussions with TUM may focus on how to adapt the annotation tools and pedagogical environment to the task of annotating statutory and regulatory texts. This includes identifying appropriate annotation types, for example, statement types such as definition or obligation, statement parts like antecedent, consequent, or exception, and agent types or roles. Moreover, given TUM’s expertise in dealing with the commercial legal sector, I will benefit from their ideas about how best to adapt annotation environments for instruction in professional business, as opposed to solely academic, contexts.

My interactions with these groups and the visits will help us to identify fruitful collaborations and to generate fundable research proposals based on this work. I expect that my DAAD-supported activities will result in a novel technologically-based component of a law school curriculum, and new techniques that law school instructors can employ to improve teaching and student learning. I also expect it will result in several conference and journal articles.

---