EasyChair Preprint

№ 9169

Toward an Intelligent Tutoring System for Argument Mining in Legal Texts

Hannes Westermann, Jaromir Savelka, Vern R. Walker, Kevin D. Ashley and Karim Benyekhlef

October 26, 2022

Toward an Intelligent Tutoring System for Argument Mining in Legal Texts

Hannes WESTERMANN [a,1], Jaromír ŠAVELKA [b], Vern R. WALKER [c], Kevin D. ASHLEY [d] and Karim BENYEKHLEF [a]

[a] *Cyberjustice Laboratory, Faculté de droit, Université de Montréal*
[b] *School of Computer Science, Carnegie Mellon University*
[c] *LLT Lab, Maurice A. Deane School of Law, Hofstra University*
[d] *School of Computing and Information, University of Pittsburgh*

Abstract. We propose an adaptive environment (CABINET) to support caselaw analysis (identifying key argument elements) based on a novel cognitive computing framework that carefully matches various machine learning (ML) capabilities to the proficiency of a user. CABINET supports law students in their learning as well as professionals in their work. The results of our experiments focused on the feasibility of the proposed framework are promising. We show that the system is capable of identifying a potential error in the analysis with very low false positives rate (2.0-3.5%), as well as of predicting the key argument element type (e.g., an issue or a holding) with a reasonably high F_1-score (0.74).

Keywords. Intelligent tutoring system, caselaw analysis, case brief, legal education, legal annotation, legal text classification, argument mining, human-computer interaction.

1. Introduction

In this paper we examine the application of cognitive computing [16] to support both a law student learning how to extract key arguments from a court opinion and a legal expert performing the same. We propose an adaptive environment that evolves from a tutoring system to a production annotation tool, as a user transitions from a learner to an expert. The concept is based on a novel cognitive computing framework where (1) the involvement of machine learning (ML) based components is carefully matched to the proficiency level of a human user; and (2) the involvement respects the limitations of the state-of-the-art of automated argument mining in legal cases. We experimentally confirm feasibility of the key ML components by testing the following two hypotheses: Given a sentence in a case brief, it is possible (H1) to detect if the sentence is placed in an *incorrect* section, and (H2) to predict the *correct* section for the sentence.

[1] Corresponding Author: Hannes Westermann, E-mail: hannes.westermann@umontreal.ca

2. Background

Lawyers routinely analyze case decisions (i.e., court opinions) to gain insight into what is a persuasive or binding precedent (typically common law countries) and/or what is the established decision-making practice in a given matter (typically civil law countries). As the list of relevant cases may be long and the opinions might be sizeable, a principled approach to the analysis is necessary to make the task feasible and as efficient/effective as possible. Such an approach requires knowing how to read an opinion, which parts to focus on, and which information to identify as crucial for understanding the case.

In U.S. law schools, case briefs are widely employed to teach law students how to analyze a case and how to use prior decisions to create new arguments or analyses [6]. Writing a case brief involves reading and understanding a case, and identifying text passages that contain the key aspects of the decision. These are then extracted and arranged in a structured format that often includes the following sections:

- **Facts** - Events and actions relevant to the dispute.
- **Issue** - Main questions (points of contention) the court must address.
- **Holding** - Legal rulings when the law is applied to a particular set of facts.
- **Procedural History** - The treatment the dispute has received from the courts.
- **Reasoning** - The analysis of the court leading to the outcome.
- **Rule** - The official rules the court must adhere to (e.g., statutory provisions).

Interestingly, many professors never ask students to turn in their briefs and, hence, do not provide a learner with much needed feedback. [18] However, practice and feedback are essential for learning. When it comes to practice, the research clearly shows that it should be focused and deliberate [8], at the appropriate level of challenge [8], and in sufficient quantity [14]. Such practice should be coordinated with targeted feedback on specific aspects of students' performance in order to promote the greatest learning gains. [2,5] Feedback should also be timely, i.e., immediate and frequent [13]. These elements do not seem present when it comes to learning to brief cases. As a result, while law students tend to start out by dutifully briefing cases, they usually switch to a less detailed approach after a few weeks, focused on color-coding sentences or taking notes in the margins of the case texts. Due to the lack of feedback and practice, it is thus unclear whether the crucial skill of briefing cases has been acquired.

To address the issue we propose CABINET, an intelligent tutoring system that gradually evolves from a platform aimed at learners to a powerful annotation environment to support an expert. In a nutshell, CABINET allows a user to select a sentence and assign it to one of the case brief's sections. More importantly, the system provides varying levels of scaffolding (i.e., varying levels of challenge) and timely feedback appropriate for the learner's level of proficiency to maximize the learning outcomes. The tool thus adapts with the user, teaching them how to brief cases at first and later supporting them in briefing and understanding cases more efficiently.

3. Related Work

Numerous researchers have proposed frameworks where a human and a computer complement each other in performing tasks in the legal domain. For example, human-aided

computer cognition framework has been proposed and evaluated in the context of eDiscovery. [15] Active learning has been explored in various contexts, such as classification of German [35] or United States [25] statutory provisions, or relevance assessment in eDiscovery [7]. The annotation tool proposed in [37] supports human annotators by enabling them to view similar sentences together. The environment described in [36] provides statistical insights into a data set assisting a human expert in creating text classification rules. The work presented in this paper is to our best knowledge the first study that explicitly maps multiple ML components to different levels of user's proficiency.

Multiple research studies have explored applications of intelligent tutoring systems in teaching legal argumentation and case analysis skills. These include supporting law students in graphically representing legal arguments [22], assessing case relevance and distinguishing cases [1], performing case-based and rule-based reasoning [4], and selecting applicable legal rules from statutes [23] and precedents [21]. An adaptive legal textbook based on knowledge graphs has been proposed in [31]. The framework presented in this paper is the first intelligent tutoring system for legal domain that can be adapted to the proficiency level of a user to eventually support a legal professional in the task of analyzing cases by extracting their key arguments.

A key component of the proposed framework is the automatic recognition of key argument elements in case texts. This task has been studied extensively in AI & Law and it is often referred to as automatic identification of rhetorical roles that sentences play in the text of courts' opinions. Rhetorical role classification focuses on segmenting cases into functional parts [39,12,28] which can, e.g., improve legal information retrieval and enable legal argument retrieval [10,33,34]. Information about a sentence's rhetorical role can also be utilized in summarization [9,11,19,3,40]. The roles often include categories such as Facts, Issue, or Conclusion that are related to the ones used in this work.

A variety of ML/NLP techniques have been employed to predict sentence role labels. These span from rule-based approaches [36] to applying ML models such as Support Vector Machines [33]. The problem has also been treated as tagging of sequences that consist of multiple sentences instead of simpler single sentence classification. Here, models such as Conditional Random Fields have often been used [24,26]. A deep learning system based on Bi-LSTM was shown to perform well in [3]. Systems based on a multilingual embeddings, Bi-LSTM, or pre-trained language models demonstrated strong transfer learning capabilities in this task [29,30]. The work presented in this paper is the first attempt to use the sentence rhetorical role identification models in intelligent tutoring to support law students in learning how to analyze legal cases.

4. Proposed Framework

We propose CABINET (CAse Brief INteractive EnvironmenT) which is a cognitive computing framework that adapts to the proficiency level of a user. An overall design of CABINET's user interface is shown in Figure 1. We adopt a fairly standard layout where an analyzed document is displayed beside a template to be populated with the extracted key argument elements. The goal is to identify the key argument spans of text (*argument element identification task*) and to categorize them in terms of case brief sections (*argument element categorization task*).

CABINET provides scaffolds and supports that allow it to evolve from an intelligent tutoring system for learners to a tool to support the more efficient work for professionals.

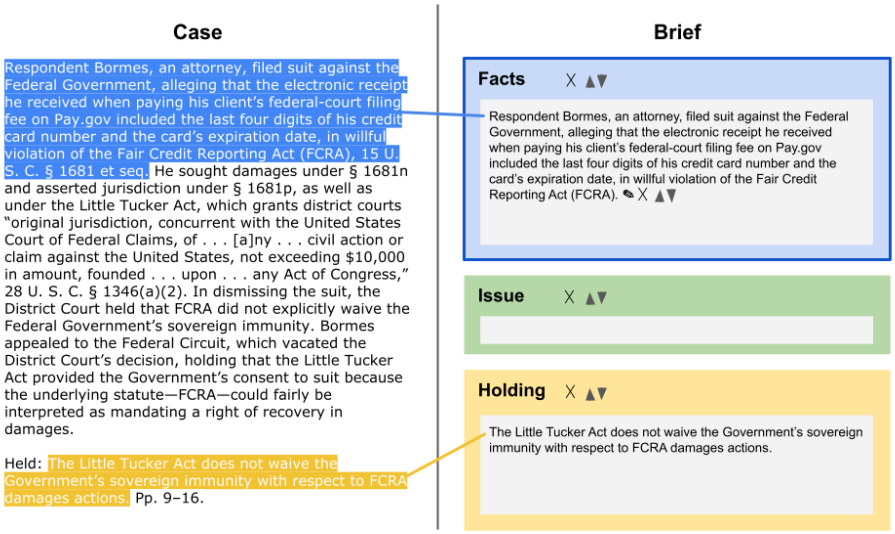

Figure 1. An overall design of CABINET's user interface: The analyzed document is displayed on the left. The case brief sections are populated and edited on the right. The system preserves the mapping between the original text and the resulting case brief sections.

To this end we adopt the National Institutes of Health's competencies proficiency scale[2] (NIH proficiency scale), a highly-regarded instrument used to measure one's ability to demonstrate competency in a task.

Figure 2 shows how the system adapts and adjusts to the proficiency level of the user as indicated by the NIH scale, from level 1 (Fundamental Awareness) to level 5 (Expert). Initially, the system provides the user with reference answers and explanations to provide an adequate level of challenge and timely feedback (blue in Figure 2). As the user learns, they are able to perform more tasks themselves (green). The system takes on the role of a safe-guard against apparent mistakes relying on one of its ML components. In the latter stages the system's ML components take over the initial steps in performing the work (orange) and a user (now an expert) reviews the results and corrects mistakes.

Fundamental Awareness (NIH level 1) - An individual at this level has common knowledge for understanding basic techniques and concepts. The learner is aware of the concept of briefing a case, its purpose and value, and is superficially aware of case brief sections. At this level, CABINET leverages the so-called *worked example effect*: studying worked examples appears to be more effective than learning by solving the equivalent problems. [32] This effect has also been confirmed in the related area of reasoning about legal cases. [20] As shown in Figure 2, the learner's task is to use the CABINET interface to inspect and reflect on cases annotated by legal education experts. The interface also provides explanations justifying the choices of the expert. This stage relies on the in-depth annotation of a small number of cases (tentatively 10-20 cases).

Novice (NIH Level 2) - At this level, an individual has the level of experience gained in a classroom and/or experimental scenarios or as a trainee on the job. They are expected to need help when performing a skill. Learners at this level are ready to attempt to categorize key argument elements with respect to case brief sections. As shown in the

[2] https://hr.nih.gov/working-nih/competencies/competencies-proficiency-scale

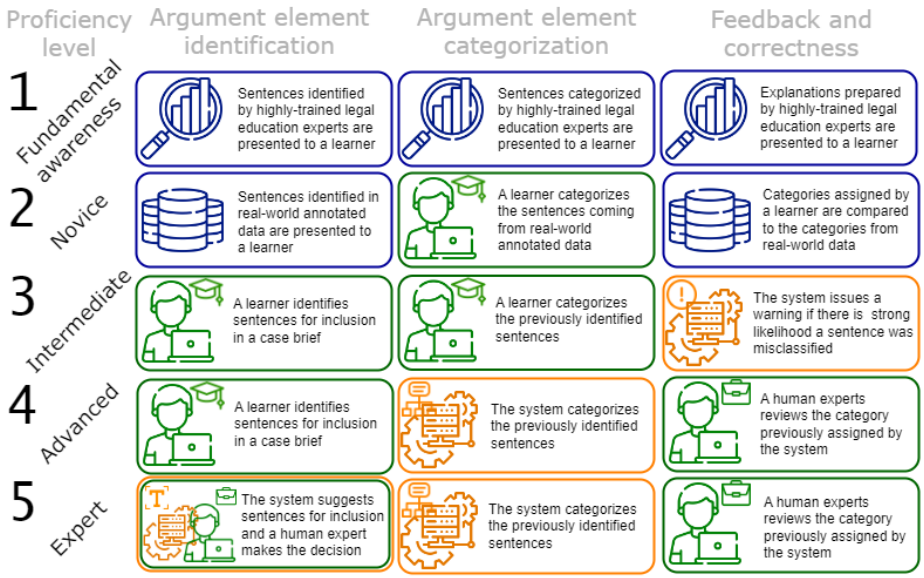

Figure 2. The rows of the diagrams correspond to the five NIH proficiency levels. The columns represent tasks performed as cooperation between a user and a computer. The performance of the tasks is either based on static expert data that have been pre-annotated (blue), user's work (green), or ML component (orange).

second row of Figure 2, learners are presented with texts where the key argument elements have already been identified by legal experts. The learner performs the argument element categorization task. Their choices are compared to those of the experts and the learner is notified about a mismatch and the category assigned by an expert is revealed.

Intermediate (NIH Level 3) - An individual at this level is able to successfully complete tasks as requested with occasional expert help. At this level, the learner attempts to identify key argument elements in a text and to categorize them. As shown in the third row of Figure 2, CABINET assumes the role of a safe-guard which evaluates the work of the advanced learner. Here, the feedback comes from a ML component that identifies the sentences that have a very high likelihood of being placed in the wrong section. The feasibility of such an ML model is evaluated in Section 5 (Experiment 1).

Advanced (NIH Level 4) - At this level, an individual can perform the actions associated with the skill without assistance. It is assumed that a trained professional can independently identify the key argument elements in a text as well as classify them with the correct case brief section. At this stage, CABINET employs a classification model to predict the case brief section of an argument element that a user identified for inclusion in the case brief. The user is expected to evaluate the predictions and correct potential errors. Some errors are tolerable at this stage, since the user is proficient enough to efficiently correct them. The feasibility of such a model is evaluated in Section 5 (Experiment 2).

Expert (NIH Level 5) - At this level, an individual is an expert in a given area. They can provide guidance, troubleshoot and answer questions related to this area of expertise and the field where the skill is used. CABINET provides the same kind of assistance as in the previous stage, but uses more active ways of supporting the professional at this level. Specifically, CABINET subtly highlights passages in a text with colors corresponding to predicted case brief sections and intensity corresponding to system's confidence in the

passage being a key argument element. This is achieved by a ML component applied to a full text. Since such predictions cannot be performed with a high degree of reliability (see Section 3), the highlighted text passages are only meant to augment the expert's review of a case text, allowing them to identify the key argument elements more efficiently.

5. Experiments

To examine the framework's feasibility, we assess two hypotheses that correspond to the system's key capabilities described in Section 4. Given a sentence in a case brief:

(H1) ... it is possible to detect if the sentence is in an *incorrect* section.
(H2) ... it is possible to predict the *correct* section for the sentence.

The capability assessed by H1 is deployed at the Intermediate proficiency level (NIH Level 3), to warn a user when a sentence is likely assigned to an incorrect case brief section. The capability assessed by H2 is utilized at the Advanced and Expert proficiency levels (NIH Levels 4 and 5), to predict the correct section for a text passage identified by a user as a key argument element.

5.1. Dataset

We obtained a dataset of 715 unique case briefs by scraping a publicly available Case Brief Summary database.[3] We used an extensive battery of regular expressions to segment the retrieved briefs into individual sections corresponding to the key argument element types. While there were over 100 unique section names we identified the six main types (see Section 2) to which we could map many of the different variations (e.g., all of "Legal Issue", "Issues", and "Issue" map to a single category). We applied a specialized legal case sentence boundary detection system to segment the sections into 9,924 sentences. [27] Figure 3 shows the distribution of the sentences in terms of the key argument element types from the perspective of their overall counts as well as their distribution over the individual case briefs. We divide the dataset into random splits on a document basis. The splits are used for training (70%), validation (15%) and testing (15%).

5.2. Model and Training

As a **baseline** we use a simple model that randomly predicts the labels based on their frequency in the dataset. We use the standard implementation provided by the sklearn framework, with "stratified" sampling.[4] As the main model we employ **RoBERTa** (a robustly optimized BERT pretraining approach) developed by Liu. [17] The model was chosen due to its high performance and simplicity to train. While higher performance might be achieved by more recent models, the RoBERTa model suffices for the purpose of proving the feasibility of the key components of the framework. Out of the available

[3] Accessible at: `http://www.casebriefsummary.com/`. Currently, the website appears to be offline. However, it has been archived by the Web Archive project at `https://web.archive.org/web/20200927234341/http://www.casebriefsummary.com/`

[4] `https://scikit-learn.org/stable/modules/generated/sklearn.dummy.DummyClassifier.html`

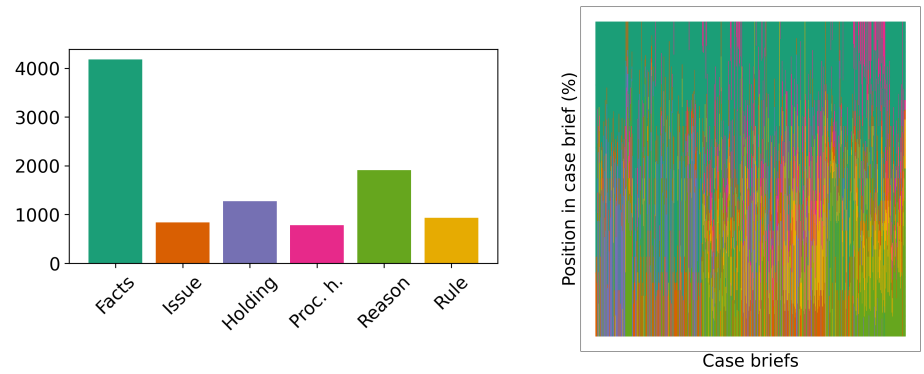

Figure 3. Left: Distribution of sentences in the dataset by section. Right: Location of sentence classes (y-axis) across the different case briefs (x-axis).

Table 1. Statistics about warning and abstentions for section assignments, for thresholds 0.05, 0.1 and 0.2.

	Warn	Abstain
Warn	5952	122
Abstain	1248	1318

	Warn	Abstain
Warn	6282	169
Abstain	918	1271

	Warn	Abstain
Warn	6549	256
Abstain	651	1184

models we chose to work with the smaller *roberta.base* that has 125 million parameters, for faster iteration times. Due to the sentences being short, we did not have to address the model's sequence length limitation of 512 tokens. The training set is used to train the model for 4 epochs. At each epoch, we evaluate the performance of the model against the evaluation set and pick the best-performing model for our experiments.

5.3. Testing H1: Warning about an Incorrectly Categorized Argument Element

When interacting with an Intermediate user (NIH Level 3) CABINET issues a warning if a piece of text identified by the user as a key argument element has been (highly likely) assigned with to an incorrect section (see Section 4). Hence, the input is a short text together with a label assigned by the user. The system either issues a warning about the assignment being likely incorrect or abstains.

We transform the dataset by creating text-label pairs between each sentence and all the labels. Since there are 1,440 sentences in the test set and 6 unique labels, there are 8,640 such pairs. For each pair, we retrieve the probability distribution the model assigns over the possible labels. If the value for a given pair is below a static threshold (we experiment with 0.05, 0.1 and 0.2), the system issues a warning. It is crucial to minimize the number of false positives (i.e., issuing a warning when the user-assigned label is in fact correct). It is comparatively less important to treat false negatives (i.e. missing out on an incorrect assignment). Since the user is still in the process of learning, abstaining in case of a mistake is preferable to providing the user erroneous (confusing) feedback.

The results for the three thresholds are reported in Table 1. The columns correspond to whether the warning should be raised, whereas the rows correspond to whether the model would raise the warning or abstain. The diagonal (cells shaded in green) reports the number of pairs for which the model behaves as desired. The cells outside of the diagonal (shaded in red) report the disagreements.

Table 2. Performance (left) and confusion matrix (right) for class predictions.

	Baseline			RoBERTa		
Argument Type	P	R	F_1	*P*	*R*	F_1
Facts	.46	.47	.47	.90	.80	.85
Issue	.05	.05	.05	.96	.91	.93
Holding	.13	.17	.15	.42	.53	.47
Procedural History	.06	.06	.06	.66	.81	.73
Reasoning	.23	.17	.20	.56	.67	.61
Rule	.06	.07	.07	.65	.48	.55
Weighted Avg	.28	.27	.27	.76	.73	.74

	Facts	Issue	Holding	Proc. H.	Reason	Rule
Facts	524	0	9	14	31	3
Issue	2	122	0	0	0	3
Holding	25	4	72	4	46	21
Proc. H.	35	3	6	95	3	1
Reason	59	4	36	5	181	38
Rule	8	1	13	0	11	61

5.4. Testing H2: Categorizing Key Argument Elements Automatically

When interacting with the Advanced and Expert users (NIH Levels 4 and 5) CABINET automatically categorizes the key argument elements identified in the text by the user (see Section 4). This is a straightforward sentence classification task. For this component, a certain amount of error is tolerable since (a) the user's proficiency is relatively high and (b) the user is actively involved in selecting the sentences. Hence, they are in a good position to verify and potentially correct and confirm the system's category assignment.

To evaluate the ability of automatically categorizing argument elements, we compare predictions of the trained RoBERTa model to the baseline. As shown in Table 2, it appears the performance differs considerably across types. Facts and Issue argument element types are identified more reliably than Holding or Rule. Table 2 shows a confusion matrix over which classes are frequently confused with other classes. The predicted labels are shown in the rows, and the true labels in the columns. For example, we can see that holdings are frequently confused with reasoning, which may be due to the small size of the classes or the classes having low "semantic homogeneity", compare [38].

6. Discussion and Future Work

The results of the experiment evaluating H1 show that the false positive rate (see Table 1) varies between 2.0% and 3.8%, depending on the threshold. This rate appears to be acceptable given the envisioned use case and user's proficiency level (Intermediate - NIH Level 3). The false negatives rate (i.e., the system abstains when a warning should have been raised) varies between 48.6% and 35.5%. While such a rate is high, we argue that in case of an isolated error due to factors such as fatigue, stress, or lack of attention, the missed warning is tolerable. If a learner has a systematic misconception, the user errors will repeat and the system will likely detect a larger portion of those. Hence, the learner will receive clear and timely feedback triggering further learning.

Evaluation of H2 shows promising performance of the fine-tuned RoBERTa model, although the performance is far from perfect. This is acceptable since this component supports a user at Advanced or Expert proficiency level (NIH Levels 4 and 5). Therefore the potential to confuse a user by an incorrect prediction is relatively low. Since the user actively selects the argument element and is immediately presented with a prediction they are in a comfortable position to perform a correction. We argue that it is far more efficient to inspect automatic predictions and make corrections when needed (in about 25% of predictions), compared to categorizing the key argument elements manually.

While the experimental results confirm our working hypotheses, there are several important limitations to the presented study. *First*, the design of the experiments only takes into account passages of text that have been selected for inclusion in the case briefs by legal experts. However, the user may occasionally make mistakes in their selections. This is particularly true for users at the Intermediary proficiency level (NIH level 3). *Second*, we do not address the challenge of assessing the current level of proficiency of the user. *Third*, the functionality of the system at the Fundamental Awareness (NIH Level 1) proficiency level requires a limited but highly curated dataset of annotated cases with detailed feedback addressing common misconceptions—a resource we have not yet created. *Fourth*, we did not conduct a feasibility study of the highlighting functionality at the Expert proficiency level (NIH Level 5). *Fifth*, and most importantly, we have not conducted a pilot user study to tentatively gauge the expected improvements in learning outcomes. We plan to address these limitations in future work.

7. Conclusion

We proposed an adaptive environment to support case law analysis based on a novel cognitive computing framework that matches various ML capabilities to the proficiency of a user. We have shown how the environment could (i) support a learner in mastering the skill of identifying key argument elements in a court opinion, and (ii) support a professional in performing the same task more efficiently. We have demonstrated that it is possible to detect if a sentence is placed in an incorrect section in case brief (H1), and to predict the actual argument element type of a case brief sentence (H2) with a reliability sufficient for the envisioned use case based on the proficiency level of a user. Hence, we have taken the initial steps in establishing the feasibility of the proposed system.

Acknowledgements Hannes Westermann and Karim Benyehklef acknowledge the generous support from the Cyberjustice Laboratory, LexUM Chair on Legal Information, and Autonomy through Cyberjustice Technologies project. Figure 2 has been designed using resources from `Flaticon.com`.

References

[1] Aleven, V. “Using background knowledge in case-based legal reasoning: a computational model and an intelligent learning environment.” *Artificial Intelligence*. v. 150, nn. 1-2, pp. 183–237. Elsevier. 2003.
[2] Ambrose, S. A. et al. *How Learning Works*, John Wiley & Sons, 2010.
[3] Bhattacharya, P., et al. ”Identification of Rhetorical Roles of Sentences in Indian Legal Judgments.” arXiv preprint arXiv:1911.05405 (2019).
[4] Bittencourt, I., Costa, E., Fonseca, B., Maia, G., and Calado, I. “Themis, a Legal Agent-based ITS.” AIED Applications in Ill-Defined Domains. p. 11. 2007.
[5] Black, P., and William, D. Assessment and classroom learning. *Assessment in Education*, 5, 7–74, 1998.
[6] Brostoff, T. and Sinsheimer, A. (2013). *United States Legal Language and Culture: An Introduction to the US Common Law System*. Ch. 3. Third Edition, Oxford University Press. 2013.
[7] Cormack, G., and M. Grossman. “Autonomy and reliability of continuous active learning for technology-assisted review.” *arXiv preprint arXiv*:1504.06868 (2015).
[8] Ericsson, K. A., Krampe, R. T., and Tescher-Romer, C. The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, pp. 363–406, 2003.
[9] Farzindar, A. & G. Lapalme. “Letsum, an automatic legal text summarizing system.” JURIX, 2004.

[10] Grabmair, M., et al. "Introducing LUIMA: an experiment in legal conceptual retrieval of vaccine injury decisions using a UIMA type system and tools." Proc. 15th Int'l Conf. on AI and Law. 2015.
[11] Hachey, B. and C. Grover. "Extractive summarisation of legal texts." AI and Law 14, 305-345, 2006.
[12] Harašta, J., et al. "Automatic Segmentation of Czech Court Decisions into Multi-Paragraph Parts." Jusletter IT 4.M (2019).
[13] Hattie, J., and Timperley, H. The power of feedback. *Rev. of Educational Research*, 77, 81–112, 2007.
[14] Healy, A. F., Clawson, D. M., and McNamara, D. S. The long-term retention of knowledge and skills. In D. L. Medin (Ed.), *The psychology of learning and motivation*, pp. 135–164, 1993.
[15] Hogan, C., R. Bauer, and D. Brassil. "Human-aided computer cognition for e-discovery." In *Proc. 12th Int'l Conf. on Artificial Intelligence and Law*, pp. 194-201. 2009.
[16] Licklider, JCR. "Man-computer symbiosis." IRE trans. on human factors in electronics 1 (1960): 4-11.
[17] Liu, Y., et al. "Roberta: A robustly optimized bert pretraining approach." arXiv preprint arXiv:1907.11692 (2019).
[18] Makdisi, M., and Makdisi, J. How to write a case brief for law school. *Introduction to the Study of Law: Cases and Materials*. 3rd Ed, LexisNexis, 2009.
[19] Moens, M.-F., "Summarizing court decisions." Info. Processing & Management 43, 6, 1748-1764. 2007.
[20] Nievelstein, F. et al. The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases. *Contemporary Educational Psychology*, 38(2), pp. 118–125, 2013.
[21] Muntjewerff, A. and Breuker, J. "Evaluating PROSA, a system to train solving legal cases." Proceedings of AIED. pp. 278–285. IOS Press Amsterdam. 2001.
[22] Pinkwart, N., Ashley, K., Lynch, C., and Aleven, V. "Evaluating an intelligent tutoring system for making legal arguments with hypotheticals." Int'l J. of AI in Education. v. 19, n. 4, 401–424. IOS Press. 2009.
[23] Routen, T. "Reusing formalisations of legislation in a tutoring system." AI Review 6, 145 – 159, 1992
[24] Saravanan, M., B. Ravindran, and S. Raman. "Improving legal document summarization using graphical models." Frontiers in Artificial Intelligence and Applications 152 (2006): 51.
[25] Šavelka, Jaromír, Gaurav Trivedi, and Kevin D. Ashley. "Applying an interactive machine learning approach to statutory analysis." In *Legal Knowledge and Information Systems*, pp. 101-110. 2015.
[26] Savelka, J., & Ashley, K. "Using conditional random fields to detect different functional types of content in decisions of U.S. courts with example application to sentence boundary detection." ASAIL 2017.
[27] Savelka, Jaromir, et al. "Sentence boundary detection in adjudicatory decisions in the united states." Traitement automatique des langues 58 (2017): 21.
[28] Savelka, J. & Ashley, K. "Segmenting US Court Decisions into Functional and Issue Specific Parts." JURIX. 2018.
[29] Savelka, Jaromir et al. "Lex Rosetta: Transfer of Predictive Models across Languages, Jurisdictions, and Legal Domains." ICAIL 2021, 129–38. https://doi.org/10.1145/3462757.3466149.
[30] Savelka, Jaromir, Hannes Westermann, and Karim Benyekhlef. "Cross-domain generalization and knowledge transfer in transformers trained on legal data." *ASAIL*, 2020.
[31] Sovrano, F., K. Ashley, P. Brusilovsky, and F. Vitali. "YAI4Edu: an Explanatory AI to Generate Interactive e-Books for Education." 4th Int'l Wkshp on Intelligent Textbooks. CEUR 3192, 31–39, 2022.
[32] Sweller, J., 2006. The worked example effect and human cognition. *Learning and instruction*.
[33] Walker, V.R., et al. "Automatic Classification of Rhetorical Roles for Sentences: Comparing Rule-Based Scripts with Machine Learning." ASAIL@ ICAIL. 2019.
[34] Walker, V. R., et al. "Semantic types for computational legal reasoning: propositional connectives and sentence roles in the veterans' claims dataset." Proc. 16th Int'l Conf. AI and Law. 2017.
[35] Waltl, B., J. Muhr, I. Glaser, G. Bonczek, E. Scepankova, and F. Matthes. "Classifying Legal Norms with Active Machine Learning." In JURIX, pp. 11-20. 2017.
[36] Westermann, H., et al. "Computer-Assisted Creation of Boolean Search Rules for Text Classification in the Legal Domain." JURIX. 2019.
[37] Westermann, H., Savelka, J., Walker, V., Ashley, K. & Benyekhlef, K. "Sentence Embeddings and High-Speed Similarity Search for Fast Computer Assisted Annotation of Legal Documents." *Jurix 2020*.
[38] Westermann, H. et al. "Data-Centric Machine Learning in the Legal Domain." ArXiv:2201.06653 [Cs].
[39] Wyner, Adam, et al. "Approaches to text mining arguments from legal cases." Semantic processing of legal texts. Springer, Berlin, Heidelberg, 2010. 60-79.
[40] Xu, H., Savelka, J., Ashley, K. "Using Argument Mining for Legal Text Summarization" JURIX. 2020.