Learning by Diagramming Supreme Court Oral Arguments

Kevin Ashley
LRDC, School of Law
University of Pittsburgh
Pittsburgh, PA, USA
ashley@pitt.edu

Niels Pinkwart
Computer Science Institute
Clausthal University of Technology
Germany
niels.pinkwart@tu-clausthal.de

Collin Lynch
Intelligent Systems Program
University of Pittsburgh
Pittsburgh, PA, USA
collinl@cs.pitt.edu

Vincent Aleven
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA, USA
aleven@cs.cmu.edu

ABSTRACT

This paper describes an intelligent tutoring system, LARGO, that helps students learn skills of legal reasoning with hypotheticals by analyzing oral arguments before the US Supreme Court. The skills involve proposing a rule-like test for deciding a case, posing hypotheticals to challenge the rule, and responding by analogizing or distinguishing the hypotheticals and/or modifying the proposed test. Using LARGO, students diagram argument transcripts in a special-purpose graphical language. LARGO provides feedback in the form of reflection questions.

Keywords

Argument models, argument diagrams, intelligent tutoring systems, hypothetical legal reasoning

1. INTRODUCTION

A problem in developing intelligent tutoring systems (ITSs) in law is the need to represent legal problems and arguments in artificial terms a tutoring system can analyze. Given problems expressed in a specially-designed representation (e.g., case frames, sets of dimensions or factors, semantic webs, or simply the responses to expert systems queries) an ITS can engage students in important aspects of legal reasoning, such as making the next move in a legal argument (see, e.g., [1,2]). Artificial representations, however, are difficult for students to employ. They must be taught the meanings of frames, factors, nodes and arcs, and the conventions for using them. If a student's resulting representations fail to satisfy quality constraints, the system may not be able to apply its expertise to help students learn. Non-uniform representations of similar facts, for instance, may foil analogically mapping problems to precedents. In any event, the artificial descriptions are not as expressive as textual ones. One might try to engineer tutoring systems to deal directly with student-authored texts, or one may try to make the process of representation serve as the primary educational vehicle.

In the work described here, we have designed an educational activity in which the process of representation *is* instructive. Students spend the bulk of the instructional time representing authentic examples of an important aspect of legal argumentation, and the system's expertise helps them to appreciate and understand the examples they represent.

ICAIL '07, June 4-8, 2007, Palo Alto, CA USA.

Our pedagogical goal is to develop law students' facility for proposing, framing, and testing legal rules using hypotheticals. Attorneys and judges reason *about* legal rules, not just with them. A proposed rule for deciding a case can be seen as a hypothesis, a tentative assumption made in order to draw out and test its normative, logical or empirical consequences. A hypothetical is an imagined situation that involves the hypothesis and helps draw out those consequences. When a student thinks of a proposed legal rule for deciding a case, can the student imagine situations in which the proposed rule would lead to unintended results or cause conflicts with applicable legal principles? Can the student use these hypothetical examples to critique the proposed rules and can she respond to such critiques?

Reasoning with hypotheticals is a conceptual tool for attorneys, judges, law professors, or students to employ in reasoning about statutory provisions, legal rules, underlying policies, and how they apply in specific situations, both in common law [9] and civil law jurisdictions [10, pp. 528-9]. For instance, in their two-page summary illustrating teleological underpinnings of the property law of wild animals as construed in *Pierson v. Post* and related cases, the Berman and Hafner posed hypotheticals, or reported that judges posed them, at least four times (e.g., suppose the quarry had been a quail rather than a fox, suppose violence among sportsmen escalated, suppose the defendant "should lie in the way with his guns, and fright the boy from going to [the competing] school, suppose the adjoining landowner uses decoys) [7].

Our LARGO ("Legal ARgument Graph Observer") ITS helps students learn skills of reasoning with hypotheticals by studying U.S. Supreme Court oral argument transcripts and representing, in a graphical format, the reasoning with hypotheticals they illustrate. Supreme Court Justices famously pose hypotheticals in oral arguments to evaluate proposed rules for deciding a case. An advocate proposes a decision rule or test. The Justices pose hypotheticals to probe the proposed test, its meaning, consistency with past decisions and underlying principles, and its legal and policy implications. The advocate responds, perhaps by distinguishing the hypothetical or modifying the proposed test.

LARGO invites students to represent the arguments in simple graphic terms. It employs a graph grammar, a set of production rules that can "parse" the diagram for purposes of drawing heuristic inferences about its completeness and sense relative to an idealized model of legal reasoning with hypotheticals. The rules are not intended to enable the program to make or respond to argument moves. Rather, they identify in the argument graphs three categories: (1) incomplete dialectical patterns, relative to the "ideal model", (2) non-standard (possibly erroneous) patterns, and (3) complete patterns under the ideal model that are worth reflecting on (and typically somewhat complex).

These dialectical patterns, when they occur in students' argument representations, offer pedagogically valuable opportunities for

reflection. LARGO offers targeted feedback on improving the representation of argument components and reflecting on their legal significance to the argument's merits.

After providing an example and model of some dialectical patterns of interpreting proposed legal tests with hypotheticals, drawn from a first year legal course, this paper discusses the use of graphic argumentation representations in ITSs for the legal domain. It then illustrates a graphic representation of a portion of the example with LARGO and explains how the system analyzes it using its graph grammar to generate a variety of feedback.

2. HYPOTHETICAL LEGAL REASONING

As an illustration of legal reasoning with hypotheticals, we have chosen a case that deals with the topic of *personal jurisdiction*, a court's power to require a person or corporation to appear in court and defend against a lawsuit. Law students encounter this technical legal concept in the first year "Legal Process" course. Casebooks focus on a set of important U.S. Supreme Court cases dealing with personal jurisdiction. The cases usually involve a court sitting in one state asserting power over a nonresident of that state and illustrate conflicts between the U.S. Constitutional principle of "due process" (i.e., minimum procedural safeguards against the arbitrary exercise of government power) and the principle that a state may redress wrongs committed within or affecting residents of the state.

In *Kathy Keeton v. Hustler Magazine,* 465 U.S. 770 (1984), the plaintiff, Kathy Keeton, sued Hustler Magazine, an Ohio corporation with its principle place of business in California, in U.S. District Court in New Hampshire. She claimed that Hustler had libeled her in five articles published in the mid 70's. Ms. Keeton was not a resident of New Hampshire and had almost no ties there. Hustler's contacts with New Hampshire included the monthly sale there of up to 15,000 issues of its *Hustler* magazine. New Hampshire was the *only* state in which Ms. Keeton's claim was not time-barred under a state statute of limitations. The District Court dismissed her claim for lack of personal jurisdiction. The Court of Appeals affirmed, concluding that she had too few ties to give the state of New Hampshire an interest in entertaining the suit, that it would be unfair for New Hampshire to assert personal jurisdiction over Hustler, given that the state's statute of limitations for libel cases was so much longer, and that, under the "single publication rule" of libel law, Hustler would potentially be liable in the New Hampshire suit for damages occurring in other states. After the lower courts rejected her suit, she appealed to the U.S. Supreme Court.

Each side in a Supreme Court oral argument gets thirty minutes to address the issues before the Court; the arguments are transcribed, recorded and later published. Figure 1 illustrates an excerpt of the *Keeton* case oral argument, in which counsel for plaintiff, Mr. Grutman, makes his arguments. The right column contains the text of the actual argument with the line numbers indicated; "Q:" indicates a Justice's question.

As Figure 1 illustrates, the advocates and Justices make a variety of standard argument moves: an advocate proposes a test for deciding the issue, the Justices pose hypotheticals to clarify or challenge the proposed test, and the advocate typically responds in a number of ways. The left column labels these argument elements. In line 14, the advocate proposes a test for deciding the problem in a manner favorable to his client and offers as a supportive reason that the test is consistent with precedent. In response, a Justice may pose a hypothetical as in lines 55, 57, and 59. The hypothetical may be a query about the test's meaning, as in lines 57 and 59. Or, the hypotheticals may underscore the (too?) broad scope and implications of the proposed rule, as in lines 55 and 59. To maintain credibility, the advocate has to rebut or otherwise reply to the challenge. As in lines 56 and 58, he may attempt to justify his test by arguing that the supposedly disanalogous hypothetical example is really analogous, disputing that a suitable rule applied to the counterexample should yield a different result. Or, he may distinguish the hypothetical from the current fact situation (cfs) and modify the test to exclude it, as in lines 64 and 66.

→ Proposed test of Mr. Grutman for Plaintiff Keeton	14. GRUTMAN: The synthesis of those cases holds that where you have purposeful conduct by a defendant directed at the forum in question and out of which conduct the cause of action arises or is generated that satisfies the formula of those minimum contacts which substantial justice and reasonable fair play make it suitable that a defendant should be hailed into that court and be amenable to suit in that jurisdiction.
← J.'s hypo	55. Q: Would it apply in Alaska?
→ Response: analogize cfs/hypo	56. GRUTMAN: It would apply, Mr. Justice Marshall, wherever the magazine was circulated. It would apply in Honolulu if the publication were circulated there. It would apply theoretically and, I think, correctly wherever the magazine was circulated, however many copies were circulated.
← J.'s hypo	57. Q: Just to clarify the point, that would be even if the plaintiff was totally unknown in the jurisdiction before the magazine was circulated?
→ Response: analogize cfs/hypo	58. GRUTMAN: I think that is correct, Mr. Stevens, so long as Alaska or Hawaii adheres, I believe, to the uniform and universal determination that the tort of libel is perpetrated wherever a defamatory falsehood is circulated. Wherever a third person reads about it, there is that harm.
← J.'s hypo	59. Q: What if the publisher had no intention of ever selling any magazines in New Hampshire?
→ Response: distinguish cfs/hypo; modify test to exclude hypo.	64, 66. GRUTMAN: It might be different. It might be different, because in that case you could not say, as you do here, that you have purposeful conduct. There you have to look for other -- I think your phrase is affiliating circumstances, other connections, judicially cognizable ties --

Figure 1. Hypothetical reasoning in *Keeton* oral argument

Legal reasoning with hypotheticals can be summarized in a more schematic form as in Figure 2 [3]. The key feature of the process is the analogical comparison of the hypothetical example to the problem situation. The reasons refer to underlying legal principles; they justify why the cfs and hypothetical are analogous or disanalogous and should be treated the same or differently. They support the criticism that the proposed test fails to achieve this result, and may guide the decision how to recover.

As Figure 1 suggests, the Justices and advocates do not always make the reasons explicit. The underlying principles that inform and justify the analogy or disanalogy may not be stated, but they seem to be a key feature. Indeed, the process of interpretive reasoning with hypotheticals serves as a short hand way of invoking the underlying principles and operationalizing a discussion of their application.

→ **Propose test and argument for deciding cfs:** Construct a proposed test that leads to a favorable decision in the cfs and is consistent with applicable legal principles and important past cases, and give reasons.
← **Pose hypothetical as probe of or counterexample to proposed test:** Construct a hypothetical example that is analogous to [disanalogous from] the cfs (i.e., a suitable test when applied to the example should yield the same [a different] result) and yet the proposed test when applied to the example leads to a different [the same] result, and give reasons.
→ **Rebut or otherwise reply to respondent's counterexample:** (1) Save the proposed test by showing that the supposedly analogous [disanalogous] example is really disanalogous [analogous]). Or (2) modify the proposed test so that it behaves like a suitable test or does not apply to the example. Or (3) abandon the proposed test.

Figure 2. Model of Interpretive Reasoning with Hypotheticals

3. REPRESENTING ARGUMENTS

Using graphic representations in learning environments to schematically represent legal arguments is not new [18, 8]. Carr [8] used Toulmin schemas [17] for collaborative legal argumentation, and the Araucaria system [15] uses premise/conclusion visual argument structures. ArguMed's argumentation "assistant" [19] provides intelligent feedback by analyzing structural relations in diagrams. By contrast, LARGO's argument representation is tailored to the dialectical pattern of proposing a test that is evaluated using hypotheticals; its ontological categories for interpreting arguments (i.e., current fact situation, tests, hypotheticals, and their relationships) help students understand complex argumentation examples. Asked to render in Toulmin schemas a transcript of oral argument before the U.S. Supreme Court, a student may do a fine job yet never realize the importance of the relationship between tests and hypotheticals.

Our special-purpose representation may be compared to [14], where domain-specific critical questions about an argument's weaknesses are represented in a general argument diagramming platform. The grain size of our features is somewhat broader than the examples provided there. While the components of our graphs (tests, hypotheticals, and modifications) are related to legal argumentation schemes identified in [20], ours represent larger argument "chunks" integrating components across a broader range of argument text. From a pedagogical viewpoint, these chunks (and the idiosyncratic and sometimes incomplete ways in which novices express them) provide good opportunities for reflection about the dialectical patterns and their significance to the argument's merits and as a method of argumentation.

LARGO allows a student to relate tests and hypotheticals to the argument transcript using simple markup techniques like Araucaria's [15]. LARGO's graph grammar, however, enables it to analyze, and provide feedback on, the student's argument representation (based on the dialectical pattern and links between the transcript and the diagram). Though promising approaches for graphically supporting argumentation exist, only Carr [8] conducted an empirical evaluation of the approach, but he was unable to show that his system caused a significant learning gain.

AI & Law researchers have used heuristics to model hypotheticals as moves along dimensions [16, 5] and related precedents and dimensions (i.e., factors) to legal principles [6], but no one has integrated these techniques to assess proposed tests by posing hypotheticals. Nor can LARGO achieve that goal, but the argument model, and techniques for representing the dialectical patterns, of hypothetical reasoning are steps in this direction.

4. LARGO ITS

LARGO engages students in reflecting upon authentic examples of legal argument as they represent them in a graphical format. The goal is to make students active learners through the representation task, instead of passive recipients as may often occur in larger law school classes.

LARGO's feedback on aspects of law student's graphical representation, generated with its graph grammar, helps it to induce students' active participation. It enables students to create graphic argument representations like the one in Figure 3 (appendix), which illustrates a depiction of the *Keeton* argument excerpts of Figure 1. (This figure was prepared by a naive user as an illustration, not by a student from LARGO's target population.) The oral argument transcript is at the left side of the LARGO screen. On the right is a workspace for creating the picture using the palette of argument representation elements at the bottom. A student can create a graph representing an argument exchange in the transcript by dragging and dropping the elements from the palette on to the workspace. Using a text highlighting feature, students can also link the elements in their graphs to passages in the transcript. LARGO's graphical representation language includes elements for representing the current fact situation, proposed tests (and modifications), hypotheticals, and various relations among them (e.g., modification of a test, distinguishing or analogizing a hypothetical, a hypothetical leading to test or modification, and a general relation). The diagram illustrates some of these including the unintentional publication hypothetical (l. 59) as distinguished from the purposeful publication in the cfs and leading to a test modification (ll. 64, 66).

At any point in the representation process, LARGO offers advice on how to improve the argument diagram by addressing three types of possible weaknesses: structural, context, and content weaknesses [12]. A s*tructural* weakness involves a part of the diagram where the relations among elements fail to correspond to the model illustrated in Figure 2. For instance, advocates commonly analogize or distinguish hypotheticals and the current fact situation; one version of a test is related to another and may have resulted from a modification proposed in response to a hypothetical. A student's graph may fail to show such relationships or may indicate uncommon relationships (e.g., analogizing or distinguishing a test and a hypothetical.) A *context* weakness occurs where the student's graph lacks elements corresponding to those in the system's marked-up version of the argument transcript. (Prior to a transcript's use in LARGO, an expert marks passages of interest, for instance, where an attorney or Justice formulates a test or poses a hypothetical.) For example, the student may have omitted a proposed test or a hypothetical, or mistaken one for the other. A *content* weakness involves a student's substandard formulation of a proposed test identified in the transcript. Given that the transcripts often do not yield explicit statements of proposed tests, formulating such a test requires interpretation: students may need to decide which conditions to include and how abstractly to characterize them, an important, though difficult, skill to learn in one's first semester in law school.

LARGO's graph grammar detects *context* and *structural* weaknesses in students' argument graphs. For detecting *content*

weaknesses, on the other hand (i.e., students' weak or un-test-like formulations of proposed tests), it employs collaborative filtering. This enables LARGO to forego natural language processing of these often intricate argument tests. Having formulated a test and enter it into a "test" diagram element, upon seeking advice, LARGO presents them with different formulations of the same test, created by other students or by their professor. LARGO asks students to select the alternative test formulations they consider of good quality. The alternatives are in themselves a form of feedback. Over time, LARGO is also able to use *each student's ranking of a small number of alternative test formulations* to produce *a comprehensive ranking of all students' formulations.* Using a collaborative filtering mechanism we developed, LARGO notifies students whose test formulations are of comparatively low quality [12].

Since a possible weakness does not necessarily mean the student has made an error, LARGO's advice is couched as suggestions for reflection. Law and legal argumentation are ill-defined domains; the system builders cannot anticipate all problems that may arise in representing arguments; a supposed "problem" may not indicate an error but an unusual argument or the fact that a Justice abandoned a line of questioning before completing a standard argument pattern. Upon detecting a possible weakness, therefore, LARGO identifies the pattern and invites the student to consider what it means and whether it should be changed.

Pattern	*Possible weakness*	*Feedback*
Incomplete pattern	No "test" element is linked to a transcript passage with a test.	Point to transcript passage; ask to represent content in the diagram.
	** Diagram contains hypothetical not related to any fact or test element.	Ask to reflect on whether the hypothetical challenges a test and what would be a good response.
Non-standard pattern	Two "test" elements connected but not by a "modified" relation.	Explain use of modified relation as normal way to connect tests
	Cfs facts and hypothetical connected but only by a general relation.	Explain analogize and distinguish relations and ask if either would be appropriate.
Complete pattern (typically complex) worthy of reflection	Diagram contains two tests not connected.	Ask to reflect on relation of the two tests.
	Diagram contains hypothetical related to cfs and test, and more than one test is present	Ask to reflect on what student's argument would be if confronted with the hypothetical.

Figure 4. Dialectical Patterns, Weaknesses, and Feedback

Examples of possible weaknesses associated with three general types of dialectical patterns LARGO's graph grammar detects are shown in Figure 4, along with the kind of feedback provided. They range from missing common elements and relations predicted by the model to non-standard connections that may indicate errors in representing an argument and completed patterns that invite reflection about their significance. Figure 3 illustrates an example of detecting a possible weakness listed in Figure 4 (marked **), a hypothetical not related to any fact or test in the diagram. LARGO suggests that the student consider whether the Alaska hypothetical (Figure 1, l. 55) should remain unrelated to any test or fact element in the diagram. The program cannot be sure that this hypothetical relates to a proposed test, but that is often so and it seems valuable to suggest that the student consider the possibility.

5. CONCLUSIONS

The LARGO intelligent tutoring system is intended to help first year law students learn legal argumentation skills by example. It engages students in studying transcripts of authentic legal arguments before the U.S. Supreme Court, and in graphically representing these arguments. The instruction focuses on dialectical patterns associated with hypothetical reasoning: proposing a test for deciding the case, posing hypotheticals to critique the test, and responding. As students develop their argument representations, LARGO provides feedback on possible weaknesses in their diagrams that are worth reflecting about, including structural, context, and content weaknesses. LARGO has been used in four unrelated legal domains.

In a controlled experiment, we assessed the utility of the LARGO tutoring approach, compared to standard mark-up techniques that law students typically use when studying legal text [13]. The experiment took place in the context of a first year, first semester Legal Process course. We designed objective pre- and post-tests to detect any differences in students' general understanding of the method and purpose of hypothetical reasoning in argument, and in their ability to recognize and reason about examples of the dialectical pattern in near and far transfer problems. We compared any learning gains of students using the LARGO system's graphical argument representations and feedback mechanisms illustrated above and of a control group addressing the same oral arguments, but using a more traditional notepad-and-highlighting environment. Consistent with our hypothesis, we found evidence that lower-ranked students (as measured by LSAT scores) (1) learned skills and knowledge about hypothetical reasoning in legal arguments significantly better than their control group counterparts for a near-transfer problem set, and (2) learned more than their control group counterparts about a central skill, evaluating hypotheticals with respect to tests.

6. ACKNOWLEDGMENTS

NSF Grant IIS-0412830, Hypothesis Formation and Testing in an Interpretive Domain, supported this work.

7. REFERENCES

[1] Aleven, V. (2006). An intelligent learning environment for case-based argumentation. *Technology, Instruction, Cognition, and Learning*.

[2] Aleven, V. (2003) "Using Background Knowledge in Case-Based Legal Reasoning: A Computational Model and an Intelligent Learning Environment." *Artificial Intelligence* 150(1-2):183-238.

[3] Ashley, K.D. (2006). Hypothesis Formation and Testing in Legal Argument. Invited paper. Inst. de Investig. Jurídicas 2d Int'l Meet. on AI and Law, UNAM, Mexico City. April.

[5] Ashley, K. (1990) *Modeling Legal Argument: Reasoning with Cases and Hypotheticals*. MIT Press/Bradford Books.

[6] Bench-Capon, T. & Sartor, G. (2003) "A Model of Legal Reasoning with Cases Incorporating Theories and Values," *Artificial Intelligence*, 150(1-2):97-143.

[7] Berman, D. & Hafner, C. (1993) "Representing Teleological Structure in Case-Based Legal Reasoning: The Missing Link" in *Proc. ICAIL 1993*. pp. 50-59. ACM Press.

[8] Carr, C. (2003) "Using Computer Supported Argument Visualization to Teach Legal Argumentation." In *Visualizing Argumentation*, 75-96. London, Springer.

[9] Eisenberg, M.A. (1988) *The Nature of the Common Law*. Harvard U. Press.

[10] MacCormick, D. N. & Summers, R.S. (ed.) (1997) *Interpreting Precedents: a Comparative Study*. Ashgate/Dartmouth.

[11] Muntjewerff, A. J., & Breuker, J. A. (2001). Evaluating PROSA, a system to train solving legal cases. In Moore, Redfield, & Johnson (Eds.), *AI-ED in the Wired and Wireless Future, Proc. AI-ED 2001* 278-285.: IOS Press.

[12] Pinkwart, N., Aleven, V., Ashley, K., & Lynch, C. (2006) Toward Legal Argument Instruction with Graph Grammars and Collaborative Filtering Techniques. In Ikeda, et al. (Eds.), *Proc. ITS 2006,* 227-236. Berlin: Springer.

[13] Pinkwart, N., Aleven, V., Ashley, K., Lynch, C. (2007) "Evaluating Legal Argument Instruction with Graphical Representations using LARGO", Proc. *AIED2007*). Marina Del Rey, CA. July

[14] Prakken, H. Reed, C., & Walton, D. (2005) "Dialogues about the burden of proof" *Proc.10 Int'l Conf. AI & Law*. ACM Press.

[15] Reed, C., & Rowe, G. (2004) "Araucaria: Software for Argument Analysis, Diagramming and Representation." *International Journal of AI Tools* 13(4):961-980.

[16] Rissland, E.L. (1989) "Dimension-based Analysis of Hypotheticals from Supreme Court Oral Argument." In *Proc. 2d Int'l Conf. on AI and Law*. 111-120 ACM Press.

[17] Toulmin, S. (1958) *The Uses of Argument*. Cambridge, Cambridge University Press.

[18] Van den Braak, S.W., Van Oostendorp, H., Prakken, H., & Vreeswijk, G. (2006) A critical review of argument visualization tools. *ECAI-06 Wkshp on Computational Models of Natural Argument.* August.

[19] Verheij, B. (2003) "Artificial Argument Assistants for Defeasible Argumentation." *Artificial Intelligence* 150(1-2):291-324.

[20] Walton, D. (2005). *Argumentation Methods for Artificial Intelligence in Law*. Berlin, Springer.

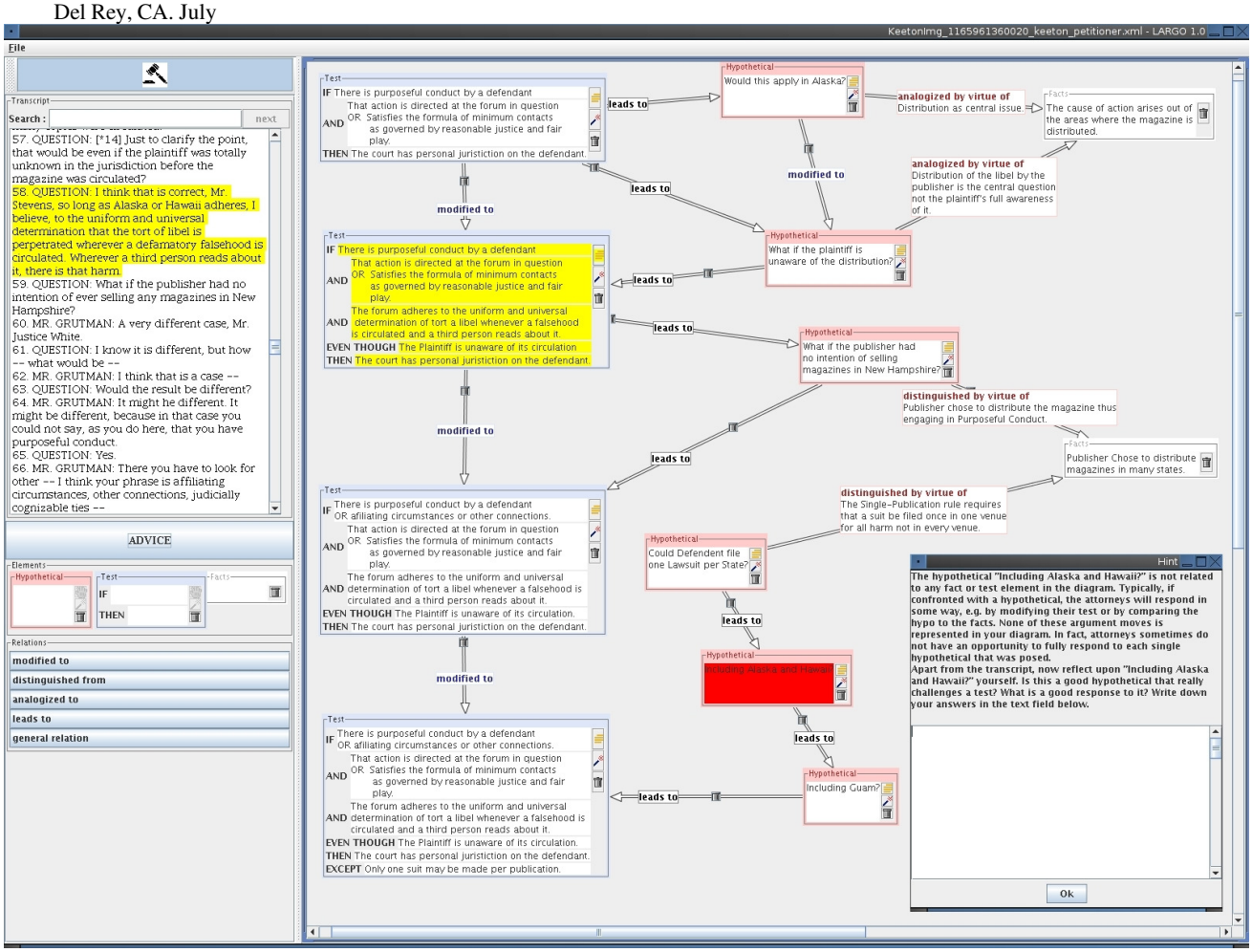

Figure 3. LARGO Representation of Keeton Case Oral Argument