T. Gross (Hrsg.): Mensch & Computer 2007: Konferenz für interaktive und kooperative Medien. München: Oldenbourg Verlag. 2007, S. 245

Student's Usage of Multiple Linked Argument Representations in LARGO

Niels Pinkwart[1], Collin Lynch[2], Kevin Ashley[3], Vincent Aleven[4]

Technische Universität Clausthal, Institut für Informatik[1]
University of Pittsburgh, Intelligent Systems Program[2]
University of Pittsburgh, School of Law[3]
Carnegie Mellon University, Human Computer Interaction Institute[4]

Abstract

The intelligent tutoring system LARGO allows law students to annotate a transcript of an oral argument using diagrams which can be linked to text portions. LARGO analyzes diagrams and gives feedback to support students' reflection. The feedback mechanisms rely on certain hypotheses about the interaction between the student and the system. In particular, a linear working mode (from beginning to end of the transcript) and a consistent and correct linking of diagram elements to the text are assumed. Based on an empirical study, this paper argues that the design of LARGO is functional and that the central interaction hypotheses are confirmed.

1 Introduction

Argumentation is of central importance in the field of law. As such, instruction in the production of reasonable arguments is a central goal of legal education and is typically taught through Socractic methods such as moot court sessions. The use of graphical representations as teaching tools has long been a common phenomenon in many domains, also in argumentation. However, although graphical representations of legal argumentation (Branting 1994) and even Supreme Court Oral Arguments (Marshall 1989) have been employed as research tools, they have not made inroads into the classroom. While legal scholars have long recognized the value of note-taking and summarization in education (Llewellyn 1930), the focus has been on textual summaries and annotation rather than on graphical markup. Multiple representations can serve several benefits in learning contexts (Ainsworth 1999), but simply providing multiple and linkable visual representations does not always improve performance – users may have problems creating links between representations. Even small changes in the way notes can be taken can have an effect on students' learning (Bauer & Koedinger 2006).

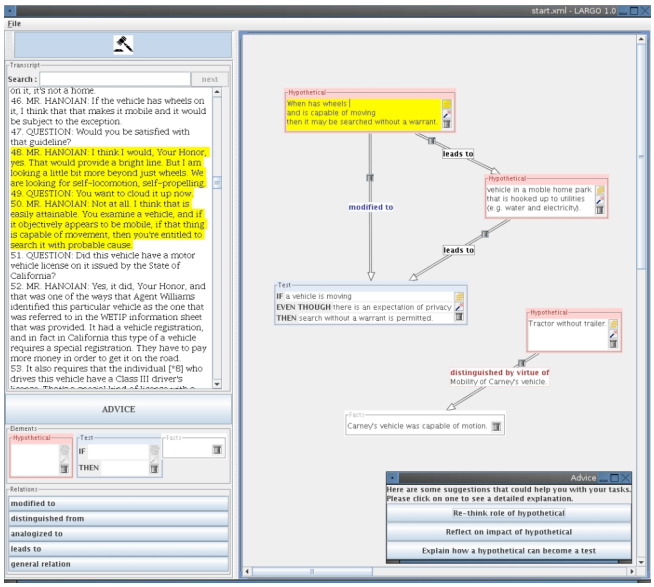

Figure 1: A partially-marked up transcript in LARGO.

This paper argues that the design of the Intelligent Tutoring System LARGO ("Legal Argument Graph Observer"), allowing students to analyze a textual argument transcript by creating graphical representations, is functional. In particular, we show that two central interaction hypotheses, which are relevant for LARGO to give advice on diagrams, are confirmed. These hypotheses are described in the next section, which starts with a brief description of LARGO and its underlying model of legal argument.

2 Argument Model and Representation in LARGO

When arguing a case at Court, advocates typically propose a *test* or legal rule that they feel should be used to decide the case. This is especially true for common-law domains such as the U.S., but also for higher levels of the roman tradition legal systems in Europe (MacCormick & Summers 1997). The justices will then often pose one or more *hypothetical cases* in order to probe the test for weaknesses. The advocate can then respond to the challenges by modifying the test, arguing that they hypothetical does not apply, or making some other change. This test-hypothetical-response interaction cycle is employed heavily during oral arguments before the U.S. Supreme Court (Ashley 1990). Students can use LARGO to study this pattern of interaction by examining and annotating transcripts of oral arguments before the Court. The system provides a suite of graphical tools that are used to summarize the action of the text by means of a specialized digraph. The graphs are composed of three node types (Tests, Hypotheticals, and Facts) and five relations (Modified-To, Distinguished-From, Analogized-To, Leads-To, and General) which correspond to the argument model. Test and

hypothetical nodes may be linked to portions of the transcript. Figure 1 is a screen shot of LARGO with a partially marked up transcript. The topmost hypothetical node ("When has wheels...") has been linked to the highlighted portion of the transcript. Once such a link has been made, the user may jump to the hyperlinked transcript portion at any time through a jump button on the node.

LARGO detects different types of *weaknesses* in argument diagrams and uses them to generate advice. *Context weaknesses* have their roots in the relationship between the graph and the text transcript. The highlighted hypothetical node in Figure 1 is an example of a context weakness. The selected portion of text contains the statement of a test, not a hypothetical, so the node's type is incorrect. The top two hints in the advice window address this. *Content weaknesses* are errors in the textual summaries within the nodes. The test node "IF a vehicle is moving..." in Figure 1 is an example of this type of weakness. The portion of the transcript that it links to should be paraphrased as "IF a vehicle is *capable of* moving...". Content weaknesses are detected via collaborative filtering, which asks students to review parts of peer's diagrams. For a more thorough discussion of LARGO see (Pinkwart et al. 2007).

Both context and content weaknesses depend on relations between the graph and transcript. The context weaknesses are *about* the relationship between the graph and the transcript. As such they cannot be detected without use of the node links. Similarly, while the content weaknesses are not about the graph linking they too depend upon it. Without some link between the graph and the transcript it would not be feasible to identify the specific element that is being paraphrased and – in turn – to have students critique peer's descriptions. Therefore, the cross-linking between the two representations is a central requirement within LARGO. Yet it is not a trivial thing to obtain. Students are free to leave the text unlinked or to select excessively large, small, or nonsensical regions. From the students perspective, the only function of the linking (quickly skipping to already-viewed areas) is not too beneficial in scenarios where the graph is viewed as a short-term squib. However it is our hypothesis (H1) that students will use the linking consistently and correctly when making their graphs.

LARGO uses the visible portion of the transcript as a clue when selecting which advice to give. In Figure 1, the transcript is displaying the text that relates to the selected node. As a result, the system chose to give hints relating to that node on the assumption that the user was examining it. This use of context is based upon several underlying assumptions. These include that the visible portion of the transcript, and the recently selected nodes, accurately indicate the user's present focus, and that the user will work forward through the transcript in a linear manner. If these are met then the feedback based upon them will be reasonably tailored. We believe that the assumptions are rational. If students are given a task of summarizing a transcript, this makes it sensible for them to read it through from beginning to end. However the argument in the transcript is complex and both parties frequently reference earlier utterances. As such the students may often need to jump back and forth in order to adequately process the references. The students are also encouraged to do so by the advice and can make use of the node hyperlinks to do so. Furthermore, the graphical notes are not linearly structured. Despite these complications we hypothesize (H2) that the students will work forward through the text incrementally reading predominantly from beginning to end.

3 Experimental Procedure and Results

In fall 2006, we conducted an experiment in order to investigate in how far LARGO can lead to better learning than a text-based alternative. The alternative tool simulates the process of examining an argument transcript with a notepad alone by allowing the students to highlight selected portions of the transcript text and enter their notes in a text pane. Details about the experimental procedure are published in (Pinkwart et al. 2007). The results with respect to the primary research hypothesis can be summarized as follows: on average, students using LARGO did better in a post-test than Control subjects, yet the difference was not statistically significant. Dividing the subjects into three groups of lower, medium, and higher ability showed that "Lower ability" subjects using LARGO scored significantly higher than their Control counterparts in several categories of important post-test items. For more skilled students, LARGO did not prove to be significantly better (but also not worse) than a notepad.

3.1 Diagram Linking

A prerequisite for LARGO's advice engine is the usage of the linking functionality: graph elements have to be linked to the parts of the transcript that they refer to. Our hypothesis H1 was that students use these functions correctly and consistently. In order to test the correctness of the linking, we chose a random sample of 15% of the diagrams created by the students and gave them to two human graders (one of them a law school professor). These graders were asked to independently judge if specific lines of the transcript, which contained hypotheticals, were represented within the argument graphs. A comparison of the two human grader's decisions showed a high inter-rater reliability (Cohen's $\kappa = 0.8$). Thus, human graders seem to agree on their decisions if a part of the transcript is represented in a diagram. We then wanted to know if LARGO's estimations about the presence of elements in diagrams are consistent with a human grader's view. This can only be possible if the diagram elements are linked to the corresponding lines of the transcript. Here, an inter-rater reliability analysis showed an agreement of $\kappa = 0.65$. We had 7 cases (out of 60) where LARGO and the human grader disagreed about whether or not some lines of the transcript, containing a hypothetical, were represented in the diagram. An analysis of these cases yielded the following results: In two cases, the hypothetical was contained in the graph, but not linked to the transcript. In two cases, the hypothetical was contained in the graph, but the reference to the transcript was to a passage where the hypothetical was mentioned *a second time* (LARGO only searched for the first appearance). In two cases, LARGO disagreed with the human grader where even the two human graders disagreed. In one case, the hypothetical formulation in the graph was too abbreviated to recognize it in a manual assessment. Overall, there was not a single case where the linked passage of the transcript did clearly not correspond to the content of the diagram element. Based on the high agreement and the subsequent analysis of the disagreement cases, we conclude that students have indeed used the linking functions *correctly*, and therefore LARGO's assessment about which elements are represented in a graph is accurate.

We then analyzed if students also used the linking mechanisms *consistently*. To do this, we counted the total numbers of elements (tests and hypotheticals) in the student's graphs, and

compared them to the numbers of elements that were linked to the transcript. The mean linking ratio was quite high (m=.87, sd=.24), which shows that the linking functionality has been used in a quite consistent manner. Was this caused by advice messages LARGO gave to students? Some of the feedback messages explicitly asked students to link graph elements to the transcript. To analyze this, we consulted the log files and counted the number of "please link" advice messages that the students read. In total, 169 of these messages were generated. This means that this type of advice was given (on average) 3 times per one-hour session. Setting the number of the messages in relation to the total number of elements in the diagrams shows that for 16% of the elements, a feedback message (asking to link this element) was shown. These numbers indicate that, while the students did to a great extent "voluntarily" use the linking mechanisms between diagram and text in LARGO, the advice messages were helpful to achieve the high link consistency that the system needs to function correctly.

3.2 Transcript Usage

The second important requirement for LARGO's feedback mechanism is that students work through the transcript in a mostly linear manner from beginning to end (H2). To analyze if this was fulfilled in the study and, consequently, the feedback produced by LARGO was adapted to the work context of the user, we consulted the log files and classified the scroll actions into four categories: long scrolls back, short scrolls back, short scrolls forward, and long scrolls forward. A scroll was counted as "long" if it moved the visible content of the transcript window by more than one page. We then compared the resulting numbers for the Experimental group with the Control group numbers. The text subjects' scrolling behavior can be used as a reasonable baseline for comparisons, since they were approaching the transcript in the same way that students approach any text when taking notes on a sheet of paper.

	LARGO (graph)	Control (text)
Long scrolls forward	0.26	0.25
Short scrolls forward	85.49	84.89
Short scrolls backward	13.51	14.59
Long scrolls backward**	0.54	0.27

Table 1: Relative scrolling frequencies

We first compared the total amounts of scrolls and found that notation medium (text vs. diagram) did not have an effect on the overall frequency of scrolling ($p>.7$). We then analyzed the relative frequencies of scroll actions for each of the four categories between the two conditions. The results confirmed our hypothesis: The groups did not differ in their behavior, and students in both groups essentially worked through the transcript in a linear manner. As Table 1 shows, the only significant difference is that students using LARGO scrolled back a long distance twice as often as the Control students. Yet, in both conditions, long scrolls back are rare. Subsequently, we analyzed in how far the difference in long scrolls back can be attributed to the use of the advice mechanism. LARGO asks students to reflect upon important parts of the transcript that are presumably not well represented – looking these up in the text might well have caused the difference in long scrolls backwards. Thus,

we consulted the log files and counted the cases where an advice message was followed by a long scroll back, either immediately or with at most one other log entry in between. This analysis indeed shows that out of 97 total long scrolls back for the Experimental condition, 55 followed an advice message. The Control condition produced 37 scrolls back in total, so that the difference between conditions can be largely explained by the advice messages. A subsequent analysis about which types of feedback messages caused the scrolls backward revealed that 52 of the 55 triggering messages indeed asked the student to re-think and reflect about not or not well represented aspects of the argument – so, the found difference between conditions represents very positive reflective activities caused by LARGO's feedback.

4 Conclusion

The analysis presented in this paper essentially confirms our research hypotheses. With some help, students used hyperlink mechanisms between text passages and self-created argument diagrams consistently and correctly, sufficient for an automatic diagram analysis based on these relations. Furthermore, we found that even when given a non-linear notation language and on-demand feedback, students still read through a textual learning resource in a linear manner. The only difference between conditions was in more long jumps back found for the graph condition. These could be explained by feedback messages asking for reflection.

References

Ainsworth, S. (1999): The functions of multiple representations. In Computers and Education (33), 131-152.

Ashley, K. D. (1990): Modeling Legal Argument: Reasoning with Cases and Hypotheticals. Cambridge MA: The MIT Press / Bradford Books.

Bauer, A., & Koedinger, K. (2006): Evaluating the Effect of Technology on Note-Taking and Learning. In CHI '06 extended abstracts, p. 520-525.

Branting, L. Karl. (1994). A Reduction-Graph Model of Ratio Decidendi. In AI and Law 2(1), 1-31.

Kozma, R.B., & Russell, J. (1996): The use of linked multiple representations to understand and solve problems in chemistry. Report Oakland University.

Llewellyn, K. N. (1930): The Bramble Bush; On Our Law and its Study. New York: Oceana Publications Inc. Dobbs Ferry.

MacCormick, D., & Summers, R. (1997): Interpreting Precedents: A Comparative Study. Aldershot, Dartmouth Publishing.

Marshall, C. C. (1989): Representing the Structure of a Legal Argument. In Proceedings of AI&Law 1989. Vancouver British Columbia: The University of British Columbia.

Pinkwart, N., Aleven, V., Ashley, K., & Lynch, C. (2007): Evaluating Legal Argument Instruction with Graphical Representations using LARGO. In Proceedings of AIED 2007.