

A Mobile Application for Self-Guided Study of Formal Reasoning

David M. Cerna

Formal Methods and Verification, johannes Kepler University
Research Institute for Symbolic Computation, johannes Kepler University
David.Cerna@risc.jku.at

Rafael P.D. Kiesel

Knowledge Based Systems, Technical University of Vienna
rafael.kiesel@tuwien.ac.at

Alexandra Dzhiganskaya

University of Applied Arts Vienna
dzhyganska.alexandra@gmail.com

In this work, we introduce **AXolotl**, a self-study aid designed to guide students through the basics of formal reasoning and term manipulation. Unlike most of the existing study aids for formal reasoning, **AXolotl** is an Android-based application with a simple touch-based interface. Part of the design goal was to minimize the possibility of user errors which distract from the learning process. Such as typos or inconsistent application of the provided rules. The system includes a zoomable proof viewer which displays the progress made so far and allows for storage of the completed proofs as a JPEG or \LaTeX file. The software is available on the google play store and comes with a small library of problems. Additional problems may be opened in **AXolotl** using a simple input language. Currently, **AXolotl** supports problems that can be solved using rules which transform a single expression into a set of expressions. This covers educational scenarios found in our first-semester introduction to logic course and helps bridge the gap between propositional and first-order reasoning. Future developments will include rewrite rules which take a set of expressions and return a set of expressions, as well as a quantified first-order extension.

1 Introduction

Logic has, over the past century, moved from an esoteric subject studied and used, in its abstract form, by the few, to a subject pervasive in the modern world. This pervasiveness is mostly due to the ubiquity of computing technology within modern society, the foundations of which rest in the realm of mathematical logic. With this in mind, one would expect formal logic to encompass a significant portion of undergraduate computer science education, however, this is unfortunately not the case [26]. Part of this problem seems to be the tremendous gap between logic studied in the abstract form and its application within computing technology.

While many can see the importance of understanding *Boolean algebra* when one wants to write a correct *if statement*, the formal theory of propositional logic seems far removed from issues like program failure or buggy software, yet it is precisely in these situations where it has helped through verification [15] and model checking [39] techniques. The software we introduce in this work does not directly provide a way to close this gap but rather focuses on getting students comfortable with formal systems as early as possible during their undergraduate education. While our work is not the first system, designed during the last few decades, with this goal in mind, to the best of our knowledge, our system is one of the first to approach self-study of formal systems using mobile technology and without limiting the system to a particular formalism, nor limiting its scope. While there exist systems such as COQ [11] and Isabelle [30], neither is particularly well suited for beginners nor can be made into a quick and easy

self-study software without significant modification. In general, such formal proof systems tend to have an extensive learning curve and are thus much better suited for advanced courses.

Our focus group consists of first and possibly second-year university students who have little or no experience with formal reasoning. Systems such as the ones mentioned in the previous paragraph are beyond their scope without significantly restricting the input language and available resources [23]. At the Johannes Kepler University (JKU), first-year students in computer science are required to take a course on logic and formal reasoning. Our software was developed to deal with educational scenarios arising from our experience over the last few years. A recent survey of the introductory logic course at the JKU [9] gave us insight into what problems the student had with the introduced material as well as what motivated them and increased their understanding. Furthermore, the results of this survey together with the development of the app lead to the introduction of a laboratory assignment during the winter semester 2019 iteration of the course which specifically uses **AXolotl**. Analysis of the effect of the application on student understanding and learning is planned for the near future.

The rest of the paper is as follows: In Section 2, we discuss existing literature and educational tools. In Section 3, we discuss rule-based formal reasoning in **AXolotl**. In Section 4, we introduce **AXolotl**, give a brief tutorial and discuss the educational scenarios it is designed to address. In Section 5, we discuss details of the implementation such as the input language and the type of problems definable in the current release. In Section 7, we discuss future work and planned extensions of the current application.

2 Related Work

Self-study assistants and educational software for logic can be separated into three categories based on the intended user interface: computer-based, web-based, and mobile-device-based. We will, for the most part, ignore physical games such as WFF’N [1]. The majority of software discovered during our literature search was either computer-based or web-based. For the most part, those tools designed for web-based interaction did not have a mobile device friendly web-design, i.e. a variant of a computer-based system. Given that the rise of mobile technology has occurred relatively recently and that the limitations of a mobile interface do not provide the optimal environment for a traditional approach to logic, we understand this dichotomy. These limitations are something we address in Section 4.

Most of the logic educational software we present below was either presented at or discussed at the following two venues: Tools for Teaching Logic (TTL) [4] and ThEdu [31]. In particular, ThEdu focuses on the use of theorem provers as educational tools, while TTL is much broader in scope. An early survey covering the existing logic education tools was published at TTL in 2011 [21]. This survey outlined over 25 different existing educational tools for logic, many of which still exist today, however, none of the tools outlined in that survey were designed for mobile devices. Rather than covering the tools outlined in [21], we will instead address trends present in the software therein discussed. For instance, the majority of the approaches taken by early developers focused on the development of software for a particular formal language [10, 17, 6]. While this is understandably a good approach within a traditional course setting, the variety of applications of logic seen today require a more flexible system, which gives the instructor the ability to freely expand and contract the formal language to a particular problem. A few existing systems taking a step in this direction are the risc program navigator [33], RISCAL [34], and Theorema [7]. The first two systems consider first-order logic (FOL) over finite models, while Theorema, defined in Mathematica, considers FOL over arbitrary models and using Mathematica’s interface restricts the student’s view of the current state of the software. Nonetheless, none of these tools has a design adequate for mobile-based use cases.

Concerning purely mobile-based software, i.e. tools developed for Android or IOS, there is much less variety in design than what can be found for other user interface categories. We conjecture that this may have to do with the steep learning curve required to write mobile applications. Furthermore, the existing mobile-based tools behave like calculators or focus on restricted fragments of the chosen formal system. These fragments are usually too weak or spurious to be of major use as a university-level educational aid. A good example of such a self-study aid is the mobile phone app *Quantifiers!* [13]. While the system is similar to our work in that it takes a fragment of the formal system and attempts to gamify it in an educationally useful way, the resulting questions tend to be of little value. Most of the quantifier examples presented tend to be contrived, i.e. they could be written much simpler without quantification. Furthermore, term construction is never needed, since the term language consists of a fixed set of constant symbols only. Given that term manipulation is one of the harder problems for students it is quite unintuitive for educational software to skip it. Other examples of similar software are *Emojic* [20], which completely abstracts logic away and focuses on image-word associations, and Lewis Carroll [35], which is written in Scratch [27] and focuses on natural language syllogisms. However, unlike the other software mentioned, Lewis Carroll does have the educational value given the role syllogisms play in philosophy education.

One mobile application that is of similar design as AXolotl is Peanoware [37] a proof construction system for natural deduction. The application has a minimalistic interface and comes with 22 built-in problems. The main goal of the app is to construct a proof for the given formula. Interestingly, the developer was able to capture the spirit of natural deduction by allowing derivations to be built both top-down and bottom-up. Furthermore, the proof construction makes use of gestures readily available on mobile devices such as drag and drop. Unfortunately, the overall design ends up impeding one's progress rather than aiding it. To give the reader an idea of the problems we are referring to, consider that no tutorial is provided with the software. While the necessary gestures are obvious, for someone new to natural deduction making the correct choice of gestures may be difficult, furthermore it may not be clear to them why a constructed proof is complete or not. At no point is one really told what they are doing.

As one may expect from a subject with its roots in philosophy, much of the existing mobile-based software is aimed at philosophy education. Note that natural deduction is a favored proof system of the subject, possibly a motivation for Peanoware. For example, the mobile app *Andor* [22] focuses on the understanding of natural language statements logically. This is one type of exercise that the interactive textbook *Carnap.io* [24] provides, though using a web-based user interface without mobile-friendly settings. Similarly, Terrance Tao developed an interactive textbook [36] for understanding the logic behind mathematical theorems. Integrating both the natural language interpretation with mathematical understanding is tackled by *Lurch*, a mathematical text typesetter [8] with an integrated prover. Though this latter software is quite out of scope, it highlights the level of development computer-based tools has reached over the past decade.

While the above outline covers many of the outliers concerning logic education software for self-study, the majority, which we have yet to mention, focus on derivation construction and proof construction assistance. A quite important example of such derivation construction tools is the *Sequent Calculus Trainer* [16] which provides a user-friendly interface for the construction of sequent calculus proofs as well as a hint engine powered by the Z3 SMT solver [29]. Note that one must use the standard sequent calculus inference rules, thus limiting the freedom of the instructor to develop more appropriate inference sets. For example, the basic logic course at the JKU uses a version of the sequent calculus better suited for an introduction to formal reasoning, i.e. rather than using sequent rules which are completely decompositional, we introduce students to rules with precise logical meaning such as *Modus Tollens*.

The already mentioned interactive textbook *carnap.io* [24] also includes a proof construction interface, but for natural deduction only. However, unlike the *Sequent Calculus Trainer* which has buttons for each rule *carnap.io* requires free-form text input from the user following a particular style of natural deduction proof representation (i.e. Fitch style, Montague style, etc.). This leaves more room for errors on the student's part, furthermore, such proof representations are not appropriate outside of philosophy education, thus limiting the applicability of the system. It is understandable why proof presentations such as Fitch style, would be used instead of the tree representation typical in computer science. Drawing a large proof tree by hand is quite a taxing endeavor. Nonetheless, there is a mismatch between these formulations of proof construction and how logic can enter the typical computer science curricula and thus making the integration of these educational tools more difficult.

Some other worthwhile mentions are the mobile app *Natural Deduction* [19] (This is a different app than Peanoware), *NaDeA: Natural Deduction assistant* [38], *The Incredible Proof Machine* [5], and *SPA: Students' Proof Assistant* [32]. Unlike most other mobile apps, *Natural Deduction* provides a natural deduction fitch style proof system as well as a theorem prover. This is, to the best of our knowledge, the most highly developed educational app for logic self-study. There exist mobile logic assistants which include theorem proving technology, for example *Logic++* [40], but they tend to put less focus on the assisting aspect and more on the proving. *NaDeA* is of similar design as *Natural Deduction* in that the student can construct proofs and have the system prove the statement for them. However, *NaDeA* can also provide hints to the students, essentially guiding them through the proof. The system is implemented on top of *Isabelle* [30] and thus benefits from its further development and expressive power. Of similar design as *NaDeA* and also implemented in *Isabelle* is *SPA* which aids students through the process of developing a proof assistant. Note that *NaDeA* highlights an earlier point we made concerning systems like *Isabelle*. They are extremely expressive but also complex to use, thus it takes work on the side of the instructor to limit the systems for first-year students.

The final self-study software we will discuss is the *Incredible Proof Machine* [5] which is a proof construction tool with a novel interface design. Rather than constructing traditional proofs, users build circuits that match the inference rules. While this is pretty standard with respect to propositional logic, it also provides an intuitive interface for more difficult calculi such as *Hilbert Systems* [28] and the *Lambda Calculus* [3]. Unfortunately, it is once again a web-based logic education tool without support for mobile devices.

Concerning our contribution to the plethora of existing software, **AXolotl**, we provide a mobile-based software, with an intuitive interface, and an expressive formal language for specifying inference rules. Essentially, the user is provided with a list of goals that they must empty using the provided inference rules. After inference application, the proof view updates so students can see their progress. So far the system can handle any variation of the standard Hilbert, sequent, and natural deduction calculi as well as encodings of the resolution and tableaux calculus within sequents. Furthermore, basic equational reasoning can be performed. However, inference rules can only have a single primary formula in their conclusion and any number of auxiliary formula. In some cases terms need to be constructed, especially for Hilbert and natural deduction calculi. When this is required, a calculator-like window opens providing an interface similar to other mobile-based applications like the *Sequent Calculus Trainer* [16], however, at no point does the student have to type the term, thus avoiding typographical errors common to other systems.

3 Rule Base Logical Reasoning

The scope of the tool lies between propositional logic and FOL, one can think of it as the quantifier-free fragment of FOL, or, equivalently as the extension of propositional logic where every atomic formula is an atomic predicate applied to terms consisting of constants and function symbols. By using the tool students shall learn something beyond a typical introduction to propositional logic, it provides preparation for reasoning in FOL. In particular, our implementation of natural deduction and Hilbert systems presents some of the less intuitive rules for beginners as a type of variable instantiation. We cover this in more detail in later sections of the paper. Before continuing, we would like to mention that by natural deduction and Hilbert systems we are not limiting our selves to particular instances of these techniques discussed below, but mainly the techniques associated with these calculi.

The logical core may be thought of as follows: we have a proof “situation” consisting of a set of “expressions” E_1, \dots, E_n where each expression is an arbitrarily nested application of “function” symbols to “constants”. In the initial proof situation, there is typically only one such expression. We have inference rules of form $\Delta, \dots \Rightarrow \Delta, \dots$, where Δ denotes an arbitrary set of expressions and \dots denotes an expression “pattern” that may contain “variables”. Note that the variables on the left and right side of \Rightarrow need not match, i.e. new variables may be introduced. On the left-side only a single pattern may occur, on the right side, there may be any number of patterns. A rule with zero patterns on the right side is an “axiom”.

An inference rule can be “applied” to a proof situation if the “ \dots ” on the left side of the rule matches (by substituting the variables by the sub-expressions) one of the expressions in the situation; the application then replaces the expression by the “ \dots ” on the right side of the rule. Here in the “ \dots ” pattern, all variables which also occur on the left side of the rule are replaced by the expressions determined by the matching of the left side; for all other variables arbitrary new expressions (e.g., chosen by the human) may be substituted. Thus, by application of an axiom, expressions are removed. The proof is complete when the situation is “empty”, no expressions left.

The didactic goal is to teach students about the basic concepts of “rules”, “patterns”, and “matching”, and furthermore the appropriate selection and application of rules matching given situations, such that by a certain number of selections and applications a certain goal is achieved.

We have experienced in our course that this thinking process is the major stumbling block for understanding the process of “first-order” reasoning; this process is independent of the particular inference rules of FOL (and other formal calculi as well). In particular, we have noted that student performance drops when these abstract notions are introduced.

In particular, one may use this framework to demonstrate “sequent style” reasoning (as done in our course) by having every expression encode a sequent; the inference rules are then (encodings of the) inference rules of the sequent calculus.

The next section demonstrates by a concrete example of how this reasoning process is implemented in the **AXolotl** software.

4 Using AXolotl

The current release of **AXolotl** requires at least API 23 of the Android operating system. This is equivalent to having release 6.0 or Marshmallow. The application is relatively small requiring only 12 megabytes of space. In order to save pictures of completed proofs, the application requires access to

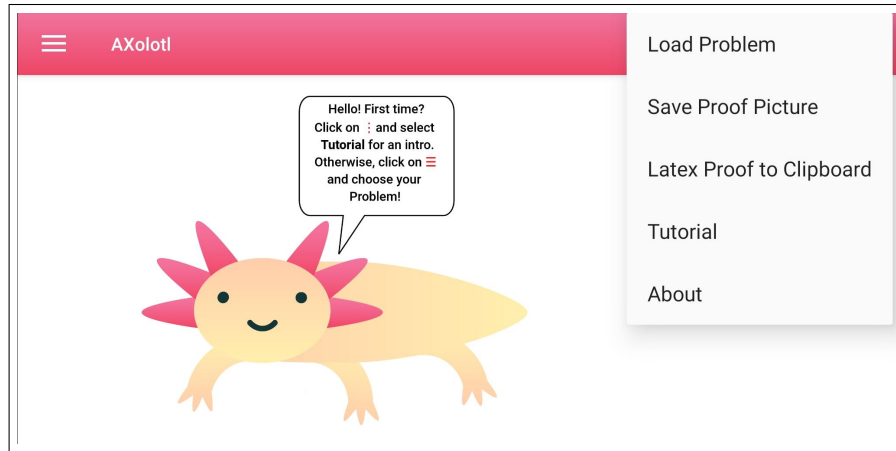


Figure 1: The options menu of AXolotl.

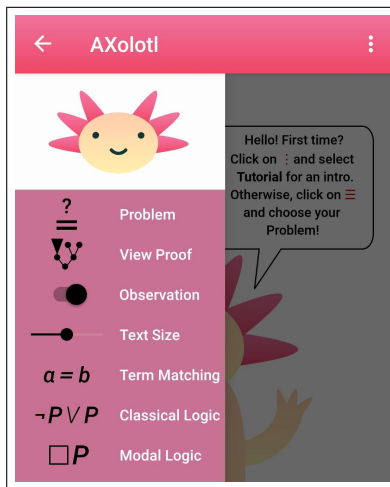


Figure 2: The drawer menu of AXolotl.

the user's internal storage directory. The application is available in the google play store¹.

When AXolotl is first opened (see Figure 1) our axolotl mascot appears providing hints concerning what to do next. Tapping on the “kebab” button opens the options menu displayed in Figure 1. If one would like to open an AXolotl file which contains a problem not found in the library then the “Load Problem” option may be used.

The following two options concern saving completed proofs: one may either save the proof as a JPEG image which may be found in the gallery of the device in the AXolotl directory or export it as a L^AT_EX file. The file is saved to the clipboard and may be copied to any other software available on the device.

The “Tutorial” option concerns first time users, it provides a short tutorial covering how to use the system. When the tutorial ends, or if one taps on the about button, information concerning the developers

¹<https://play.google.com/store/apps/dev?id=6871709124320468307>

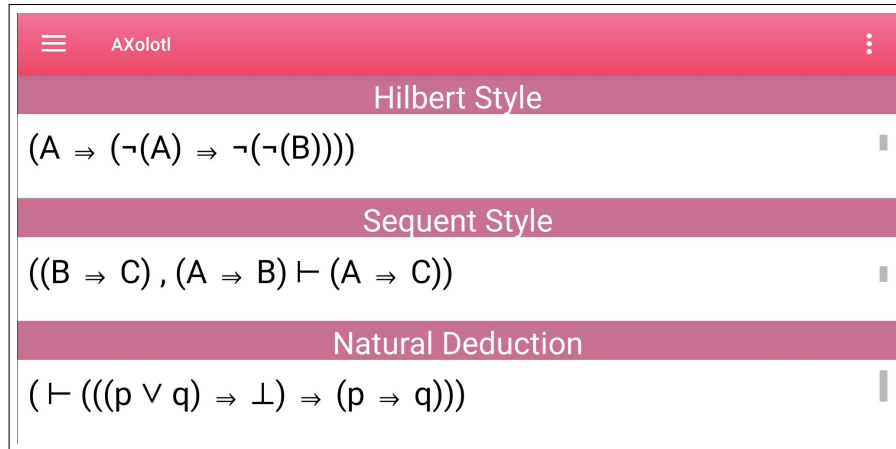


Figure 3: Problem library for classical logic problems.

is provided.

If instead of opening the “Options” menu one taps on the “hamburger” button, the drawer menu opens as seen in Figure 2. The first two buttons of the drawer menu concern the currently loaded problem, of which there currently is none.

Concerning “Observation”, AXolotl provides users with a step-by-step application of inference rules that demonstrates how an expression is changed by the chosen inference rule.

This can be quite verbose and can slow down the pace of the user. When a user is advanced enough “Observation” can be turned off providing a much smoother and quicker experience.

Concerning “Text Size”, when expressions get large, using the standard font size may result in much-unwanted scrolling. To avoid this, we allow the user to adjust the font size to be between 10 and 50 SP (Android’s unit for screen invariant font size). The rest of the menus concern the built-in libraries of problems. Future upgrades of the system will contain extensions of these libraries.

Let us now consider the library for classical logic, see Figure 3. Notice that there are three categories of problems, namely Hilbert style, sequent style, and natural deduction. As one might guess, these categories denote the style of deduction used to solve the problems. As was done in the tutorial available within the app, we will showcase the software using sequent style deduction and transitivity of implication (the problem visible in Figure 3). A proof of this property would be written as follows using Gentzen’s sequent calculus:

$$\frac{\frac{\frac{A \vdash A}{(B \rightarrow C), A \vdash A, C} \text{weak}}{(A \rightarrow B), (B \rightarrow C), A \vdash C} \rightarrow: l}{(A \rightarrow B), (B \rightarrow C) \vdash (A \rightarrow C)} \rightarrow: r$$

By tapping on the problem a new display will appear containing the selected problem, the rules and an image of our mascot (See Figure 4). Note that the rules are displayed in a succinct form which, for more complex rules, may be hard to read. Before moving on, we would like to discuss our use of Δ in the inference rules. If there is more than a single goal in the goal display, any additional goals are abbreviated by Δ . Consider Δ as an additional context that will be ignored during the application of an inference rule.

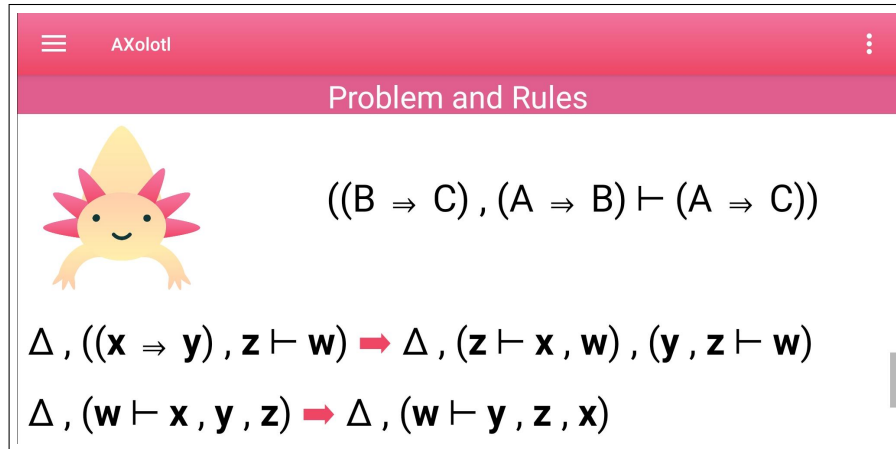


Figure 4: Problem display.

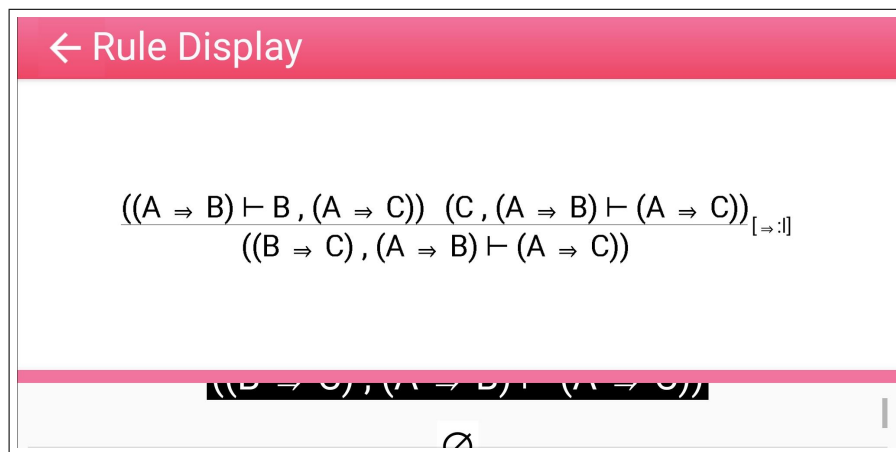


Figure 5: Rule display.

If one long taps on a rule, a new display opens which shows a pretty printed form of the rule (see Figure 5). Notice that the rule display allows you to select an \emptyset . This displays the rule without instantiating the formula variables. The rule, as displayed in Figure 5, is instantiated by our selected problem.

Once a rule and a problem have been selected in the display shown in Figure 4, they are highlighted black, one may swipe right using a fling motion to proceed to the next screen. When *observation* mode is activated AXolotl will present the user with the individual instantiation of the formula variables.

In Figure 6, the instantiation of the formula variable w is displayed. The top portion of the display “terms to match” indicates which part of the rule is matched to which part of the goal. The lower portion of the display “matching substitution” shows the precise substitution of w . Upon swiping left using a fling motion, the instantiation of the next variable will be displayed. Once all four instantiations have been displayed AXolotl switches to the screen displayed in Figure 7.

Note that at any point one can swipe right to left with a fling motion to go back to the previous display. On the problem display screen, this motion will undo the previous rule application unless there are no rule applications to undo.

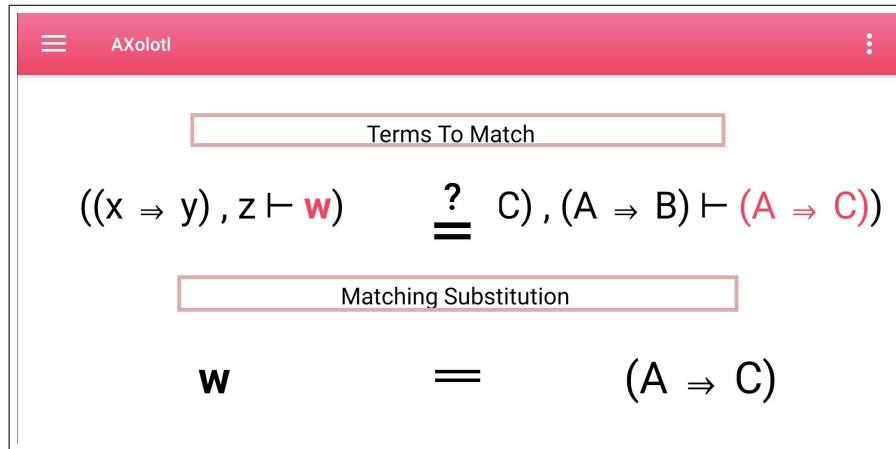


Figure 6: Instantiation of the variable w .

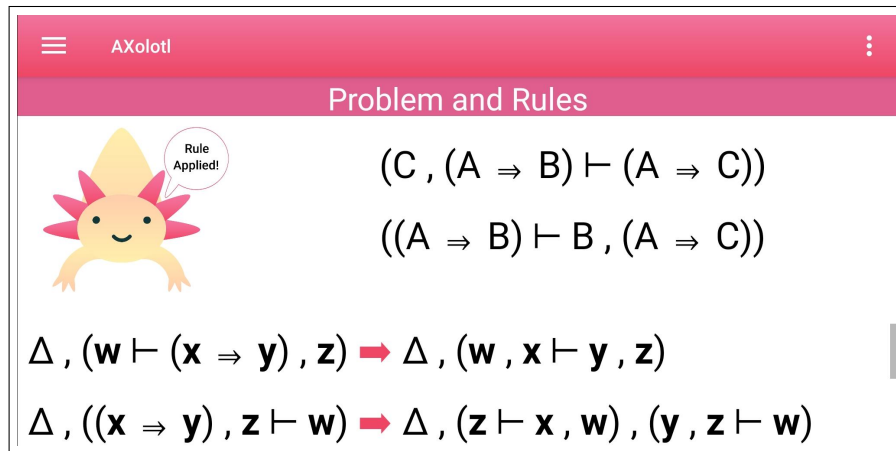


Figure 7: Problem display after rule application.

Rules

$$\frac{(z \vdash x, w) \quad (y, z \vdash w)}{((x \rightarrow y), z \vdash w)}_{[\rightarrow :]} \quad \frac{(y, z, x \vdash w)}{(x, y, z \vdash w)}_{[\text{shift} :]}$$

$$\frac{(w, x \vdash y, z)}{(w \vdash (x \rightarrow y), z)}_{[\rightarrow :r]} \quad \frac{}{(x, y \vdash x, z)}_{[\text{AX}]}$$

$$\frac{(w \vdash y, z, x)}{(w \vdash x, y, z)}_{[\text{shift} :r]}$$

Proof

$$\frac{? \quad ?}{((A \rightarrow B) \vdash B, (A \rightarrow C)) \quad (C, (A \rightarrow B) \vdash (A \rightarrow C))}_{[\rightarrow :]} \quad \frac{}{(B \rightarrow C), (A \rightarrow B) \vdash (A \rightarrow C)}$$

Figure 8: The proof view after one rule application.

$$\frac{\frac{(A \vdash A, B, C)}{((A \rightarrow B), A \vdash B, C)}_{[\text{AX}]} \quad \frac{(B, A \vdash B, C)}{((A \rightarrow B), A \vdash B, C)}_{[\rightarrow :]} \quad \frac{(A, C \vdash A, C)}{((A \rightarrow B), A, C \vdash C)}_{[\text{AX}]} \quad \frac{\frac{(C, B, A \vdash C)}{(A, C, B \vdash C)}_{[\text{shift} :]} \quad \frac{(B, A, C \vdash C)}{(B, A, C \vdash C)}_{[\rightarrow :]}}{(A, C \vdash A, C)}_{[\text{AX}]} \quad \frac{(C, (A \rightarrow B), A \vdash C)}{((A \rightarrow B), A, C \vdash C)}_{[\text{shift} :]}}{(B \rightarrow C), (A \rightarrow B), A \vdash C)}_{[\text{shift} :]}$$

$$\frac{(A, (B \rightarrow C), (A \rightarrow B) \vdash C)}{((B \rightarrow C), (A \rightarrow B) \vdash (A \rightarrow C))}_{[\rightarrow :r]} \quad \frac{((B \rightarrow C), (A \rightarrow B) \vdash (A \rightarrow C))}{((A \rightarrow B) \vdash ((B \rightarrow C) \Rightarrow (A \rightarrow C)))}_{[\rightarrow :r]} \quad \frac{((A \rightarrow B) \vdash ((B \rightarrow C) \Rightarrow (A \rightarrow C)))}{(\vdash ((A \rightarrow B) \Rightarrow ((B \rightarrow C) \Rightarrow (A \rightarrow C))))}_{[\rightarrow :r]}$$

Figure 9: The proof after the exhaustive application of inference rules to the problem.

Now if we tap on the “hamburger” menu and select “View Proof” the progress we have made so far is displayed (see Figure 8). The proof view is split into two parts, the rule list and the proof. The rule list contains all the allowed rules for the given problem. The proof contains the current state of the proof. Notice that the two branches end in question marks. This means that the two branches are currently open, i.e. do not end with axioms.

If we were to apply inference rules exhaustively the resulting proof would be as displayed in (see Figure 9). When an inference rule is applied which empties the goal set an animation is displayed ending with the image seen in Figure 10.

The proof can also be exported as a latex file with page size set to A2. The rules are not included in the latex output, only the proof.

Note that all inference rules used in this example do not introduce new variables. This is typical for sequent calculus style proofs, however, the fundamental deduction rule of Hilbert systems breaks this assumption, that is modus ponens:

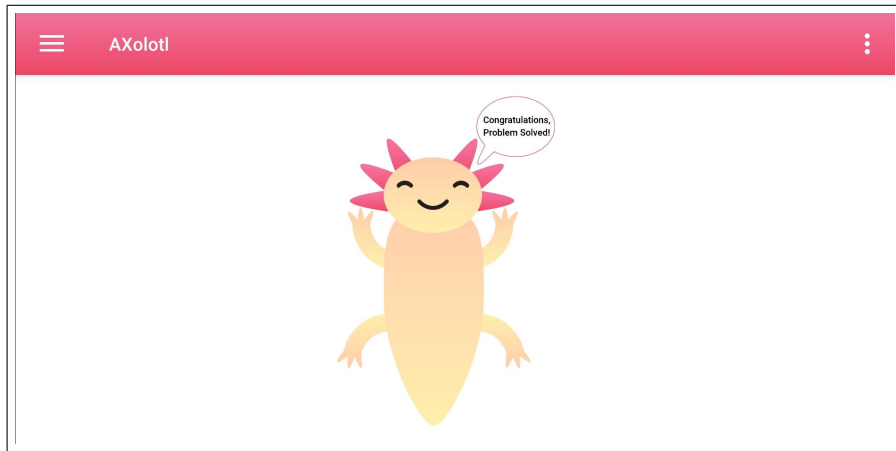


Figure 10: Displayed when goal list is emptied.

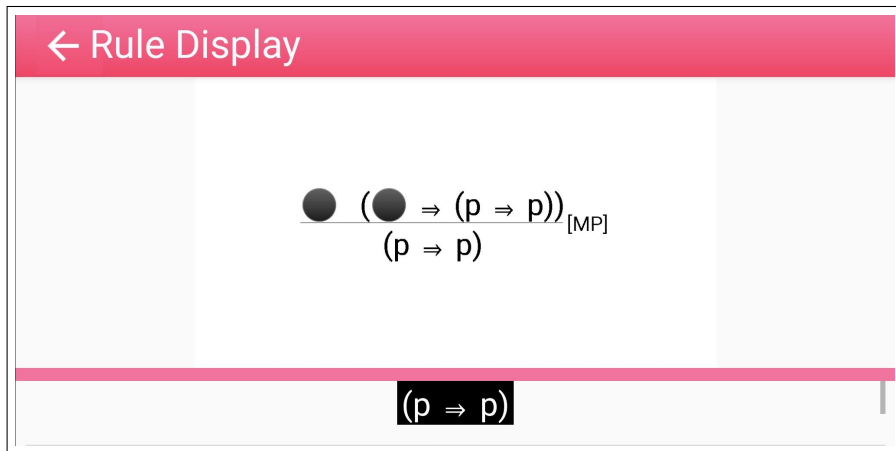


Figure 11: A rule with a hole to be filled by the user.

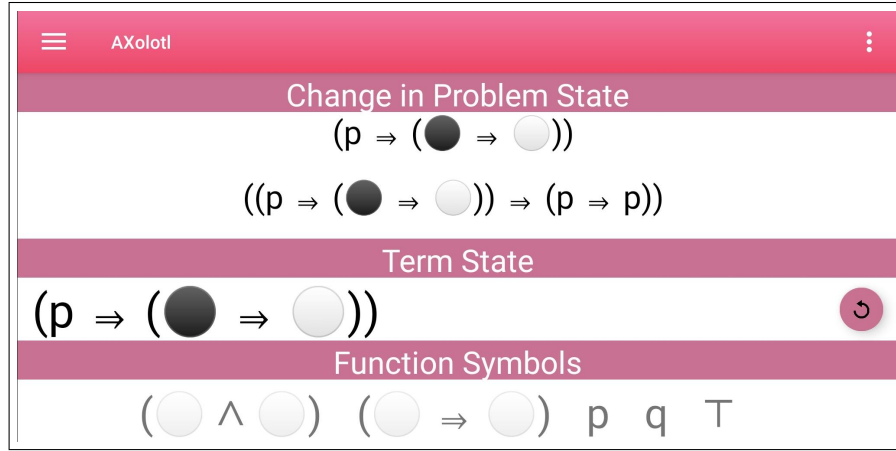


Figure 12: The term construct display allows the user to construct terms from the allowed function symbols.

$$\frac{x \quad x \Rightarrow y}{y} \text{MP}$$

The conclusion of the rule does not contain the variable x . This means that **AXolotl** requires input from the user specifying what term is to replace x .

Consider proving $P \Rightarrow P$ using modus ponens and the axioms: $x \Rightarrow (y \Rightarrow x)$ and $(x \Rightarrow (y \Rightarrow z)) \rightarrow ((x \Rightarrow y) \Rightarrow (x \Rightarrow z))$.

If we select the inference rule modus ponens the rule display will contain the contents of Figure 11. The hole visible in Figure 11 denotes the value which must be feed to **AXolotl**. If we swipe right and attempt to apply this rule we will eventually come across the screen displayed in Figure 12 which can be used to construct terms.

Note that turning off “Observation” does not mean that Figure 12 can be skipped, **AXolotl** does not provide instantiations of variables whose value cannot be derived from context. There are two types of holes displayed in Figure 12, The selected hole and the unselected holes. There may be only one selected hole at a time but there may be any number of unselected holes. Once there are no holes of either type present one may swipe right using a fling motion. If at any point one swipes left the constructed term is erased and one must start from scratch.

The “Change in Problem State” portion of the display shows what the goals list will look like based on your current instantiation of the variable. The “Term State” displays the current term you have constructed, and the “Function Symbols” portion is a list of allowed function symbols. If a function symbol takes arguments each argument will be denoted by an unselected hole. After choosing a function symbol with holes, the left-most unselected hole will be switched to selected. Notice that on the right side of “Term State” is an additional button for undoing the previous instantiation.

This covers the basic functionality and usage of **AXolotl**. The last thing we would like to cover in this section is how to use **AXolotl** for non-logical problems, for instance, show that two distinct permutations of a list are equivalent modulo associativity and commutativity, see Figure 13.

Most of the inference rules are used to transform the goal. The last inference rule is instead used to remove a particular term from the goal list. Essentially this is a variation of the equation:

$$a \circ b \circ c \circ d \equiv d \circ c \circ b \circ a$$

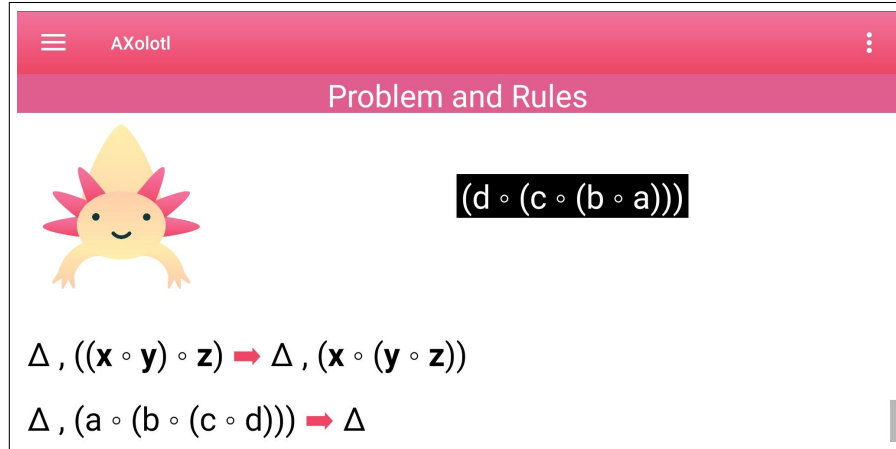


Figure 13: A rewriting problem in AXolotl.

many other problem instances, such as linear and binary search can also be written within AXolotl using a similar encoding.

5 AXolotl Problem Specification

The problems found in the built-in library are loaded on start-up from so-called AXolotl files. These files are written in a simple language usable by an instructor or savvy student to design their own problems. In Figure 14 one finds the AXolotl file for the first problem we discussed in Section 4.

The files are read line by line and there should not be any trailing line after the final input line. Each line has a prefix which may be one of the four values: “Function”, “Variable”, “Problem”, “Rule”. All “Function” and “Variable” lines must occur before “Problem” and “Rule” lines because syntactic checks are run on the “Problem” and “Rule” statements based on the allowed symbols. There are no order restrictions between “Variable” and “Function” nor between “Problem” and “Rule”. The “Function” lines must contain a symbol which may be alphanumeric and an arity (number of arguments). If arity is set to zero then the function is considered a constant. Each function symbol must have a unique arity. Additionally, a binary function can have infix as the last statement on the line implying that it ought to be printed infix rather than prefix.

“Variable” lines may only contain an alphanumeric symbol. In future releases of the software, there may be further indicators added concerning which type of variable one is defining. This would be necessary for extensions of this software to first-order where a distinction between formula variables and term variables would need to be made.

The “Problem” line starts with a number indicating how many goals will be contained in the problem’s goal set. In Figure 14 there is precisely 1 goal. At this point, one may notice that the “Problem” line of Figure 14 contains additional function symbols that were not defined in the file. These function symbols are predefined and have semantic meaning within the system. They are as follows:

- \vdash denotes sequents and must contain two lists as arguments.
- $cons(\cdot, \cdot)$ denotes the list constructor.
- ε denotes the empty list.

```

Function:  $\Rightarrow$  2 infix
Function:  $\neg$  1
Function: p 0
Variable: x
...
Problem:  $1 \vdash (\varepsilon, \Rightarrow(p, \neg(\neg(p))))$ 
Rule:  $1 \vdash (\text{cons}(x,w), \text{cons}(y,z)) \vdash (w, \text{cons}(\Rightarrow(x,y),z))$  [ $\Rightarrow:r$ ]
Rule:  $2 \vdash (z, \text{cons}(x,w)) \vdash (\text{cons}(y,z), w) \vdash (\text{cons}(\Rightarrow(x,y),z), w)$  [ $\Rightarrow:l$ ]
Rule:  $1 \vdash (\text{cons}(x,w), z) \vdash (w, \text{cons}(\neg(x),z))$  [ $\neg:r$ ]
Rule:  $1 \vdash (w, \text{cons}(x,z)) \vdash (\text{cons}(\neg(x),w), z)$  [ $\neg:l$ ]
Rule:  $1 \vdash (w, \text{cons}(\text{cons}(y,z),x)) \vdash (w, \text{cons}(x, \text{cons}(y,z)))$  [shift:r]
Rule:  $1 \vdash (\text{cons}(\text{cons}(y,z),x), w) \vdash (\text{cons}(x, \text{cons}(y,z)), w)$  [shift:l]
Rule:  $0 \vdash (\text{cons}(x,y), \text{cons}(x,z))$  [AX]

```

Figure 14: A rewriting problem in **AXolotl**

We assume that any non-list term contained in a list cannot contain a list within it, that any singleton term t , which is not a variable, contained within a sequent is a list of the form $\text{cons}(t, \varepsilon)$, and that the sequent symbol does not occur within a list of terms or within any term for that matter. There are a few additional assumptions concerning printing, but they do not affect the logical interpretation of the problem.

Before we move on, the last constraint concerning the problem line is that variables cannot occur. This will most likely change in future iterations of the software.

The “Rule” line starts with a number indicating the number of premises the rule has. The second rule in Figure 14 has two premises while the last rule has none. Rules without premises are axioms and close branches. There is no limit to the number of premises, but when outputting the proof to \LaTeX rules are restricted to five.

The next n terms are the premises of the rule where n is the number of premises. The $n + 1$ term denotes the conclusion of the rule. Every rule must have a conclusion. Additionally, to the premises and conclusion, a rule may be given a name that is displayed next to it in the rule and proof display. This name comes last on the rule line and is to be written in square brackets. It is not required, but advisable.

6 **AXolotl** in the classroom

In the winter semester 2019, we introduced **AXolotl** into our introductory logic course as part of an optional laboratory assignment. Students usually complete the laboratory assignment when they need extra credit given their performance on previous in-class exams or when they are interested in topics that go beyond the required materials. Typically 30-40% of the students partake in the assignment. This optional laboratory assignment is presented to the students between the modules covering propositional logic and first-order logic thus perfectly fitting the educational scenario **AXolotl** was designed for. In particular, we used **AXolotl** as an educational aid for introducing the students to Natural Deduction. Note that Natural Deduction tends to be problematic for students being that it is easy for students to

make errors in rule application and secondly, some aspects of rule application are quite close to first-order reasoning. For instance elimination of implication. Note that the Curry-Howard isomorphism between simple type theory and intuitionistic Natural Deduction maps implication introduction and elimination rules to the same λ -terms as universal introduction and elimination, see Chapter 3 [18].

The goal of the introduced Laboratory assignment was to get students use to thinking in a way closer to what is required for first-order reasoning. As already mentioned in the previous paragraph, the standard Natural Deduction calculus, unlike the sequent calculus or resolution, contains inference rules which correspond to the quantification inferences of first-order logic. Though this correspondence is far from equivalence, we believe it provides enough insight into the subject to help students through what seems to be the most difficult part of the course [9]. While we have not, as of yet, experimentally validated the tacitly mentioned hypothesis outlined above, we are currently working on the development of experimental scenarios that can test our hypotheses.

Concerning the excerpt from the laboratory assignment provided in Figure 15 and how it relates to the above discussion one only needs to consider how the proofs of these various statements look.

Notice that it not only requires the use of implication elimination but also a contradiction. The combination of these two rules provides a quite good simulation of the relationship between unification and quantification of first-order logic. While this type of exercise can be done without the aid of **AXolotl** it would require students to be more attentive when applying inference rules something which may be hard for beginners who are prone to typographical and consistency errors.

The rest of the questions from the Laboratory excerpt concern similar issues but using different inference rule sets and different assumptions. Given the flexibility of **AXolotl**'s input language, we can easily switch between rule sets while maintaining a user-friendly environment.

7 Future work

In previous sections, we have briefly mentioned a few points which we plan to tackle in the near future. One of the most important developments will be extending the input language as well as the system to allow rules with multiple conclusions.

Simple and important calculi such as resolution [25] and tableaux [12] cannot be expressed in a natural way using our rule descriptions; rather than the proof situation being stored as a set of expressions, as in the **AXolotl** implementation of the sequent calculus, it is stored within a single expression.

Furthermore, our proof display is inadequate for displaying the intermediary steps within a resolution proof, this would require drawing a forest of trees. Also, this additional flexibility will allow users to express even more calculi than the current input language allows.

A simpler yet more challenging change would be the addition of quantifiers. This would require minimal changes to the look and feel and input language but would require extensive reworking of the current unification and matching algorithms [2]. For those who are familiar with higher-order reasoning, we would need to perform matching on the pattern fragment of higher-order logic [14].

Educationally, we plan to perform several experiments concerning the usage of the software in the classroom, as a self-study aid, and as a tool for bridging propositional and first-order reasoning. The application was included in the winter semester 2019 iteration of our introduction to logic course as a laboratory assignment with this particular use case in mind. Analysis of the effect of the application on student understanding and learning is planned for the near future.

As a final remark, while the current problem library is of reasonable size, we plan to continuously update the problem list. We are always looking for new problems to add to the library as well as new

...

2) (.5 points) Download the file *ContrapositiveProblem.txt* which contains the problem $\vdash (\neg q \rightarrow \neg p) \rightarrow (p \rightarrow q)$ together with the appropriate rules. Load the file into AXolotl and solve. Save an image of the completed proof.

3) (1 points)

a) Replace the line in *ContrapositiveProblem.txt*
 Problem: $1 \vdash (\varepsilon, \Rightarrow (\Rightarrow (\neg(q), \neg(p)), \Rightarrow (p, q)))$
 by the line
 Problem: $1 \vdash (\vee (\neg(q), q), \Rightarrow (\Rightarrow (\neg(q), \neg(p)), \Rightarrow (p, q)))$
 Remove the line
 Rule: $1 \vdash (\text{cons}(\neg(x), z), \perp) \vdash (z, x) \quad [\perp]$
 and add the following lines
 Rule: $1 \vdash (z, x) \vdash (z, \vee(x, y)) \quad [\vee: I1]$
 Rule: $1 \vdash (z, y) \vdash (z, \vee(x, y)) \quad [\vee: I2]$
 Rule: $3 \vdash (\text{cons}(x, z), w) \vdash (\text{cons}(y, z), w) \vdash (z, \vee(x, y)) \vdash (z, w) \quad [\vee: E]$
 If you have problems doing the necessary changes yourself you can download *ContrapositiveProblem2.txt*. Using this new set of rules prove that the new problem statement is indeed provable. Save an image of the completed proof.

b) For those of you who are not familiar with the statement $(\neg q \Rightarrow \neg p) \Rightarrow (p \Rightarrow q)$ it is referred to as the *contrapositivity axiom* and can be proven equivalent to the *principle of excluded middle* $\neg p \vee p$. Proving $(\neg q \rightarrow \neg p) \rightarrow (p \rightarrow q) \vdash (\neg q \vee q)$ using the sequent calculus is quite simple, is it provable using the rule set of *ContrapositiveProblem2.txt* or is some assumption missing? If an assumption is missing what is it? (see the file *ContrapositiveProblem3.txt*)

c) What about adding
 Rule: $1 \vdash (\text{cons}(\neg(x), z), \perp) \vdash (z, x) \quad [\perp]$
 to *ContrapositiveProblem3.txt*, does it change the situation? Can be found in *ContrapositiveProblem4.txt*.

d) What is the relationship between contradiction and the principle of excluded middle? Note that contradiction may be written as $(\neg(p) \Rightarrow \perp) \Rightarrow p$. Prove the following statement

$$(\neg p \vee p) \vdash (\neg(p) \Rightarrow \perp) \Rightarrow p$$

using the rules in the file *ContrapositiveProblem5.txt*. Save an image of the completed proof.

Figure 15: Excerpt from Lab assignment.

Implication within the propositional natural deduction framework can be thought of as a meta-level quantification over formula. While the Hilbert system we use for classical propositional logic shares these properties as well given that the only inference rule, *Modus Ponens* is essentially implication elimination, it lacks the intuitive feeling of Natural Deduction and thus will most likely leave students frustrated. Even for the initiated, this feeling of frustration is well known and can still be felt.

| Rules | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------|
| $\frac{(w \vdash y) \quad (w \vdash (y \Rightarrow x))}{(w \vdash x)}_{[\Rightarrow:E]}$ | $\frac{(\neg(x), z \vdash \perp)}{(z \vdash x)}_{[Contra]}$ |
| $\frac{(z \vdash (x \Rightarrow \perp))}{(z \vdash \neg(x))}_{[\neg:df1]}$ | $\frac{(y, z, x \vdash w)}{(x, y, z \vdash w)}_{[shift:I]}$ |
| $\frac{(z \vdash \neg(x))}{(z \vdash (x \Rightarrow \perp))}_{[\neg:df2]}$ | $\frac{(x, z \vdash y)}{(z \vdash (x \Rightarrow y))}_{[\Rightarrow:I]}$ |
| $\frac{}{(x, y \vdash x, z)}_{[AX]}$ | |
| $\frac{(z \vdash \perp)}{(z \vdash x)}_{[L:E]}$ | |
| Proof | |
| $\frac{\frac{\frac{(\neg(q), p, (\neg(q) \Rightarrow \neg(p))) \vdash \neg(q))}{(\neg(q), p, (\neg(q) \Rightarrow \neg(p))) \vdash \neg(p))}_{[AX]} \quad \frac{\frac{((\neg(q) \Rightarrow \neg(p)), \neg(q), p \vdash (\neg(q) \Rightarrow \neg(p)))}{(p, (\neg(q) \Rightarrow \neg(p)), \neg(q) \vdash (\neg(q) \Rightarrow \neg(p)))}_{[shift:I]} \quad \frac{((\neg(q) \Rightarrow \neg(p)), \neg(q), p \vdash (\neg(q) \Rightarrow \neg(p)))}{(\neg(q), p, (\neg(q) \Rightarrow \neg(p))) \vdash (\neg(q) \Rightarrow \neg(p)))}_{[AX]}}{(\neg(q), p, (\neg(q) \Rightarrow \neg(p))) \vdash \neg(p))}_{[\neg:df2]}}{(\neg(q), p, (\neg(q) \Rightarrow \neg(p))) \vdash (p \Rightarrow \perp)}_{[\Rightarrow:E]}}{(\neg(q), p, (\neg(q) \Rightarrow \neg(p))) \vdash \perp}_{[Contra]}}{(\neg(q), p, (\neg(q) \Rightarrow \neg(p))) \vdash q}_{[\Rightarrow:I]}}{((\neg(q) \Rightarrow \neg(p)) \vdash (p \Rightarrow q))}_{[\Rightarrow:I]}}{\vdash ((\neg(q) \Rightarrow \neg(p)) \Rightarrow (p \Rightarrow q))}_{[\Rightarrow:E]}$ | |

Figure 16: AXolotl proof of contrapositivity.

categories. The current list reflects necessary problem types for testing the software.

References

- [1] Layman E. Allen (1966): *WFF 'N PROOF : the game of modern logic*. Autotelic Instructional Materials Publishers.
- [2] Franz Baader & Tobias Nipkow (1998): *Term Rewriting and All That*. Cambridge University Press, New York, NY, USA, doi:10.1017/CBO9781139172752.
- [3] Hendrik Pieter Barendregt, Wil Dekkers & Richard Statman (2013): *Lambda Calculus with Types*. Perspectives in logic, Cambridge University Press, doi:10.1017/CBO9781139032636. Available at <http://www.cambridge.org/de/academic/subjects/mathematics/logic-categories-and-sets/lambda-calculus-types>.
- [4] Patrick Blackburn, Hans van Ditmarsch, María Manzano & Fernando Soler-Toscano, editors (2011): *Tools for Teaching Logic - Third International Congress, TICTTL 2011, Salamanca, Spain, June 1-4, 2011. Proceedings*. Lecture Notes in Computer Science 6680, Springer, doi:10.1007/978-3-642-21350-2.
- [5] Joachim Breitner (2016): *Visual Theorem Proving with the Incredible Proof Machine*. In Jasmin Christian Blanchette & Stephan Merz, editors: *Interactive Theorem Proving*, Springer International Publishing, Cham, pp. 123–139, doi:10.1007/978-3-319-43144-4_8.
- [6] K. Broda, J. Ma, G. Sinnadurai & A. Summers (2006): *Friendly e-tutor for Natural Deduction*. In: *BCS-FACS, Practice and Experience*, doi:10.5555/2228206.2228208.
- [7] Bruno Buchberger, Tudor Jebelean, Temur Kutsia, Alexander Maletzky & Wolfgang Windsteiger (2016): *Theorema 2.0: Computer-Assisted Natural-Style Mathematics*. *Journal of Formalized Reasoning* 9(1), pp. 149–185, doi:10.6092/issn.1972-5787/4568. Available at <https://jfr.unibo.it/article/view/4568>.

- [8] Nathan C. Carter & Kenneth G. Monks (2014): *Using the Proof-Checking Word Processor Lurch to Teach Proof-Writing*.
- [9] David M. Cerna (2019): *Evaluation of the VL Logic (342.208-9) 2018W End of Semester Questionnaire*. RISC Report Series.
- [10] Dan Christensen (2010): *DC Proof*. Need: Available at <https://www.dcproof.com/>.
- [11] Thierry Coquand & Gérard Huet. (2019): *The Coq Proof Assistant*. Available at <https://coq.inria.fr/>.
- [12] Marcello D'Agostino (1999): *Tableau Methods for Classical Propositional Logic*, pp. 45–123. Springer Netherlands, Dordrecht, doi:10.1007/978-94-017-1754-0_2.
- [13] Spoon Developers (2018): *QUANTIFIERS! - A Mathematical Logic Game*. Available at <https://play.google.com/store/apps/details?id=spoon.quantifiers>.
- [14] Gilles Dowek (2001): In Alan Robinson & Andrei Voronkov, editors: *Handbook of Automated Reasoning*, chapter Higher-order Unification and Matching, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, pp. 1009–1062, doi:10.1016/B978-044450813-3/50018-7. Available at <http://dl.acm.org/citation.cfm?id=778522.778525>.
- [15] V. D'Silva, D. Kroening & G. Weissenbacher (2008): *A Survey of Automated Techniques for Formal Software Verification*. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 27(7), pp. 1165–1178, doi:10.1109/TCAD.2008.923410.
- [16] Arno Ehle, Norbert Hundeshagen & Martin Lange (2017): *The Sequent Calculus Trainer with Automated Reasoning - Helping Students to Find Proofs*. In: *Proceedings 6th International Workshop on Theorem proving components for Educational software, ThEdu@CADE 2017, Gothenburg, Sweden, 6 Aug 2017.*, pp. 19–37, doi:10.4204/EPTCS.267.2.
- [17] Olivier Gasquet, François Schwarzentruher & Martin Strecker (2011): *Panda: A Proof Assistant in Natural Deduction for All. A Gentzen Style Proof Assistant for Undergraduate Students*. In Patrick Blackburn, Hans van Ditmarsch, María Manzano & Fernando Soler-Toscano, editors: *Tools for Teaching Logic*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 85–92, doi:10.1093/jigpal/jzm026.
- [18] Jean-Yves Girard, Paul Taylor & Yves Lafont (1989): *Proofs and Types*. Cambridge University Press, New York, NY, USA.
- [19] Jukka Häkkinen (2018): *Natural Deduction. Proof generator. Proof checker*. Available at <http://naturaldeduction.org/>.
- [20] Stefan Haustein (2017): *Emojic*. Available at <https://play.google.com/store/apps/details?id=org.flowgrid.emojic>. Repository location: <https://github.com/stefanhaustein/EmojiC>.
- [21] Antonia Huertas (2011): *Ten Years of Computer-based Tutors for Teaching Logic 2000-2010: Lessons Learned*. In: *Proceedings of the Third International Congress Conference on Tools for Teaching Logic, TICTTL'11*, Springer-Verlag, Berlin, Heidelberg, pp. 131–140, doi:10.1093/jigpal/jzm019. Available at <http://dl.acm.org/citation.cfm?id=2021573.2021589>.
- [22] Matthias Jenny (2018): *Andor: Learn Logic*. Available at <https://play.google.com/store/apps/details?id=io.cordova.andor>.
- [23] Maria Knobelsdorf, Christiane Frede, Sebastian Böhne & Christoph Kreitz (2017): *Theorem Provers as a Learning Tool in Theory of Computation*. In: *Proceedings of the 2017 ACM Conference on International Computing Education Research, ICER 2017, Tacoma, WA, USA, August 18-20, 2017*, pp. 83–92, doi:10.1145/3105726.3106184.
- [24] Graham Leach-Krouse (2017): *Carnap: An Open Framework for Formal Reasoning in the Browser*. In Quaresma & Neuper [31], pp. 70–88, doi:10.4204/EPTCS.267.5. Available at <http://arxiv.org/abs/1803.00722>.
- [25] Alexander Leitsch (1997): *The Resolution Calculus*. Springer-Verlag, Berlin, Heidelberg, doi:10.1007/978-3-642-60605-2.

- [26] J. A. Makowsky & A. Zamansky (2017): *Keeping Logic in the Trivium of Computer Science: A Teaching Perspective*. *Form. Methods Syst. Des.* 51(2), pp. 419–430, doi:10.1007/s10703-017-0301-z.
- [27] Majed Marji (2014): *Learn to Program with Scratch: A Visual Introduction to Programming with Games, Art, Science, and Math*. No Starch Press.
- [28] J.D. Monk (1976): *Mathematical Logic*. *Graduate Texts in Mathematics*, 37, doi:10.1007/978-1-4684-9452-5.
- [29] Leonardo de Moura & Nikolaj Bjørner (2008): *Z3: An Efficient SMT Solver*. In C. R. Ramakrishnan & Jakob Rehof, editors: *Tools and Algorithms for the Construction and Analysis of Systems*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 337–340, doi:10.1109/MS.2006.117.
- [30] Tobias Nipkow, Lawrence C. Paulson & Markus Wenzel (2002): *Isabelle/HOL - A Proof Assistant for Higher-Order Logic*. *Lecture Notes in Computer Science* 2283, Springer, doi:10.1007/3-540-45949-9.
- [31] Pedro Quaresma & Walther Neuper, editors (2018): *Proceedings 6th International Workshop on Theorem proving components for Educational software, ThEdu@CADE 2017, Gothenburg, Sweden, 6 Aug 2017*. *EPTCS* 267. Available at <http://arxiv.org/abs/1803.00722>.
- [32] Anders Schlichtkrull, Jørgen Villadsen & Andreas Halkjær From (2019): *Students' Proof Assistant (SPA)*. In Pedro Quaresma & Walther Neuper, editors: *Proceedings 7th International Workshop on Theorem proving components for Educational software*, Oxford, United Kingdom, 18 July 2018, *Electronic Proceedings in Theoretical Computer Science* 290, Open Publishing Association, pp. 1–13, doi:10.4204/EPTCS.290.1.
- [33] Wolfgang Schreiner (2011): *Computer-Assisted Program Reasoning Based on a Relational Semantics of Programs*. In: *Proceedings First Workshop on CTP Components for Educational Software, THedu'11, Wroclaw, Poland, 31th July 2011.*, pp. 124–142, doi:10.4204/EPTCS.79.8.
- [34] Wolfgang Schreiner (2018): *Theorem and Algorithm Checking for Courses on Logic and Formal Methods*. In: *Proceedings 7th International Workshop on Theorem proving components for Educational software, THedu@FLoC 2018, Oxford, United Kingdom, 18 July 2018.*, pp. 56–75, doi:10.4204/EPTCS.290.5.
- [35] Terry Tao (2012): *Lewis Carroll*. Available at <https://scratch.mit.edu/projects/2486639/>.
- [36] Terry Tao (2018): *QED*. Available at <https://teorth.github.io/QED/>.
- [37] Laurent Théry: *Peanoware-Natural Deduction*. Google Play.
- [38] Jørgen Villadsen, Andreas Halkjær From & Anders Schlichtkrull (2019): *Natural Deduction Assistant (NaDeA)*. In Pedro Quaresma & Walther Neuper, editors: *Proceedings 7th International Workshop on Theorem proving components for Educational software*, Oxford, United Kingdom, 18 July 2018, *Electronic Proceedings in Theoretical Computer Science* 290, Open Publishing Association, pp. 14–29, doi:10.4204/EPTCS.290.2.
- [39] Willem Visser, Klaus Havelund, Guillaume Brat, SeungJoon Park & Flavio Lerda (2003): *Model Checking Programs*. *Automated Software Engineering* 10(2), pp. 203–232, doi:10.1023/A:1022920129859.
- [40] Yale Weiss (2016): *LOGIC++*. Available at <https://yaleweiss.common.gc.cuny.edu/logic/>.