Formal Modeling (WS 2022)
Some Proseminar Topics

Wolfgang Schreiner
Research Institute for Symbolic Computation (RISC)
Wolfgang.Schreiner@risc.jku.at

Your task is to study a particular mathematical/computational problem from literature, to develop a formal model of this problem in the RISCAL language, and to validate this model with the RISCAL software. The model may be the specification of a computational problem by a pair of pre/post-conditions; these may be validated by analyzing the inputs/output pairs allowed by the specification and by checking the correctness of an algorithm with respect to this specification. The model may also describe a computational system by an initial state condition and a transition relation; these may be validated by analyzing the sequences of system states arising from this model. Your goal is to find a way to formalize the model and to validate this formalization that is adequate for the chosen problem.

The following topics are only examples, you may also suggest other ones (to be approved by the lecturer). Most topics may also be pursed in the frame of *bachelor theses* (in this case, also algorithms have to be formulated and appropriately annotated such that verification conditions can be generated and checked).

Topic: Formalization of Floating Point Arithmetic

A floating point number is essentially a triple (s, m, e) of sign $s = \pm 1$, mantissa $m \in \mathbb{N}$ with $n < 2^n$, and exponent $e \in \mathbb{N}$, for some $n \in \mathbb{N}$; it describes the rational number $x = s \cdot (1 + m/2^n) \cdot 2^e$. All s, m, e are represented as sequences of bits; the IEEE 754 standard defines the details of the representation including special values such as ±0, $\pm\infty$ and NaN (not a number).

The goal is to model (an appropriately chosen subset of) this standard and some arithmetic operation (e.g., floating point addition) such that from given inputs in binary form the appropriate output is appropriately determined.

Topic: Formalization of Coding Theory

In communications and information processing, a code is a reversable mapping of strings of symbols to other strings (encoding/decoding). The goal is to formalize some part of coding theory, e.g., error-detecting codes, error-correcting codes (ECC), lossless data compression techniques (runlength-encoding, Lempel Ziff compression), etc. This involves the formalization of basic notions, code formats, potentially also the modeling of corresponding algorithms. For this, the appropriate literature has to be investigated, you may start by investigating the Wikipedia pages on these topics

Topic: Formalization of the Knuth-Morris-Pratt Algorithm

The Knuth-Morris-Pratt (KMP) algorithm solves the problem of searching in a string s for (the first position respectively all positions) of a string w. While the naive searching algorithm has asymptotic time complexity $|s| \cdot |w|$, the KMP algorithm has complexity $|s| + |w|$. The algorithm proceeds in two phases: in the first phase w is analyzed and translated into a table W; the second phase uses W rather than w for searching in s. The goal is to model the problem of the precomputation of W and the problem of the search in s with respect to W (both by suitable pre/postconditions).

Topic: Formalization of a Data Structure

Formalize some computer science data structure, e.g., some variant of a (self-balancing) binary search tree, a hash table, a disjoint set structure, or similar. The goal is to model a small theory of the data structure with the basic notions and operations; with the corresponding operations; for this, the suitable literature has to be investigated (you may start by investigating the Wikipedia notions of above topics). The formalization may be based on simple mathematical data types, e.g., a pointer-linked data structure may be represented as an index into an array whose values are the nodes of the structure.

Topic: Formalization of Binary Decision Diagrams

A binary decision diagram is a directed acyclic graph (representing a binary tree with sharing of subtrees) that encodes a multi-ary boolean function. In a nutshell, the leaves of the diagram are constant truth values while every non-leaf node represents a boolean variable; its outgoing edges denote those subtrees that describe the value of the function for the two possible values of the variable. The goal is to model the basic notion of binary decision diagrams and their evaluation and modeling the computation of binary decision diagrams from propositional formulas.

Topic: Formalization of First-Order Logic

The manuscript "Concrete Abstractions" contains a formalization of propositional logic. Along similar lines, develop a formalization of first-order logic with universal and existential quantification. Model the syntax and semantics of first order logic, basic notions such as "free variables", the computation of normal forms of first-order formulas, etc.

Topic: Formalization of a Puzzle/Game

Choose some puzzle or (e.g., card/tabletop) game and model it as a non-deterministic state transition system with some initial state, a set of actions, and a termination (winning/losing) condition. Formulate also some invariants that every state of the game must satisfy and check in RISCAL whether the game indeed preserves these invariants.