CISC-499 Projects 2018–19

Kai Salomaa
School of Computing, Queen's University
ksalomaa [AT] cs [DOT] queensu [DOT] ca

1. **Computational complexity of decision problems for regular languages**

   This is a theoretical topic requiring an understanding of basic notions in computational complexity and familiarity with time/space bounded reducibility. The topic would be suitable for students who are taking CISC-462 in fall 2018.

   It is known that all natural problems, like membership, emptiness, equivalence etc. are decidable for finite state automata. However, what is the complexity of these problems?

   The goal of this project is to investigate what are the known complexity results for the basic decision problems for deterministic and nondeterministic finite automata and for regular expressions. Often the questions are complete for PSPACE for NFAs and regular expressions, or log-space complete for DFAs. In particular, goals of the project include to identify

   - examples of natural problems for finite automata/regular expressions where the precise complexity is unknown,
   - examples finite automaton problems that are not known to be solvable, or that are known to be unsolvable.

   The goal of the project is to present the findings in the report in a systematic way (the terminology in different articles appearing in literature may not always be consistent). The project involves a fairly *large amount of literature search* since the complexity results are not included in typical textbooks. I can provide some survey articles to be used as a starting point.

2. **Converting finite-state machines to regular expressions**

   The well known state elimination algorithm converts a (nondeterministic) finite automaton (NFA) to a regular expression. This algorithm tends to generate very large and redundant regular expressions, partly because it does no simplification. In this project you will develop and implement heuristics for simplifying regular expressions and apply these heuristics to the problem of converting finite state machines to regular expressions. The tasks can include the following:

   - study the expressions that tend to be generated in conversion of finite state machines, and characterize the types of simplification that would be useful
   - develop and implement a simplification heuristics for regular expressions

- test the efficacy of your algorithms by using them in the conversion of finite state machines to regular expressions
- investigate some of the current literature on simplification of regular expressions
- (optional, if time permits) evaluate the cost of the simplification heuristics and develop a metric for deciding when to use them

The algorithm converting NFAs to regular expressions has been implemented in various libraries such as

- Fado `http://fado.dcc.fc.up.pt/`, or,
- Vaucanson `http://vaucanson-project.org/?eng`

The software libraries provide a collection of operations to manipulate finite-state machines and convert them to regular expressions or vice versa.

The project requires an understanding of the basics of finite automata and regular expressions. The amount of programming required is not large, but you should expect to run a significant number of simulations and other experiments.

3. **Average complexity of the subset construction**

   It is well known that in the worst case the minimal deterministic finite automaton (DFA) equivalent to a given n state nondeterministic finite automaton (NFA) needs 2^n states. On the other hand, the determinization algorithm based on the subset construction often works well in practice.

   The goal of this project is to study the *average case complexity of the NFA-to-DFA transformation* (strictly speaking, the NFA-to-minimized DFA transformation) by running the subset construction algorithm on a large number of randomly generated NFAs and minimizing the resulting DFAs. Determinization of NFAs and DFA minimization are included in software libraries such as *Fado, Vaucanson,* or the older package *Grail.* For information on the software libraries see topic 2. above.

   In addition to running experiments on randomly generated NFAs the goals of the project include:

   - identifying different types of NFAs where the determinization cost is close to the worst case,
   - searching in the literature for theoretical work dealing with average case complexity of the subset construction algorithm.

Other projects. I have other possible theoretical topics – please come to talk with me in my office or send me email. If you have your own idea for a project related to my research, please come to talk with me about it.