

**MAP-i Doctoral Program in Computer Science**  
**Thesis Proposal 2012**

## **Topic**

Average-case representational and operational complexity of regular languages

## **Supervisors**

- Rogério Reis ([rvr@dcc.fc.up.pt](mailto:rvr@dcc.fc.up.pt)), Departamento de Ciência de Computadores, FCUP
- Nelma Moreira ([nam@dcc.fc.up.pt](mailto:nam@dcc.fc.up.pt)), Departamento de Ciência de Computadores, FCUP

## **Research Unit**

CMUP – Centro de Matemática da Universidade do Porto  
(Web:[cmup.fc.up.pt](http://cmup.fc.up.pt))

## **Brief Description**

Regular languages play a fundamental part in Computer Science and were intensively studied during the 1960s, and have again, in the last years, become a matter of fruitful research, both at the theoretical and the applied level. This renewed interest is explained by the important role that new applications of automata theory play in fields such as computational linguistics, bioinformatics, speech and image recognition, software certification and computer networks, to mention only a few. One of the areas that has received special attention is the descriptive complexity of formal languages, which includes complexity measures, relative conciseness, complexity of language operations and of conversions between representations. These studies are motivated by the need to have good estimates of the amount of resources required to manipulate those representations, in particular the conversion between different representations of regular languages, such as finite automata and regular expressions. This is crucial in new applied areas where automata and other formal models are used, for instance for security certificates in formal verification systems. In general, having smaller objects improves our control on software, which may become shorter, more efficient and certifiable. Worst-case analysis for most classical algorithms are well known. The practical performance of these algorithms, however, is sometimes surprising. The best performing algorithms are not necessarily the ones with the best worst-case complexity. This motivates complexity analyses in the average-case, where enumeration and random generation of representations are essential.

The aim of this proposal is to contribute to the average-case studies of regular languages manipulations. Those include conversions between equivalent representations and regularity preserving operations using different models. In particular, the main topics to address from an average-case point of view, are:

- conversions of regular expressions (RE) to small non-deterministic automata (NFA's) and vice-versa;
- fine-grained analysis of the power set construction;
- operational state complexity of basic regularity preserving operations

## References

- [1] Sabine Broda, António Machiavelo, Nelma Moreira, and Rogério Reis. On the average state complexity of partial derivative automata. *International Journal of Foundations of Computer Science*, 22(7):1593–1606, 2011.
- [2] Sabine Broda, António Machiavelo, Nelma Moreira, and Rogério Reis. On the average size of Glushkov and partial derivative automata. *International Journal of Foundations of Computer Science*, 23(5):969–984, 2012.
- [3] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. CUP, 2008.
- [4] Markus Holzer and Martin Kutrib. Nondeterministic finite automata - recent results on the descriptional and computational complexity. *Int. J. Found. Comput. Sci.*, 20(4):563–580, 2009.
- [5] Markus Holzer and Martin Kutrib. Descriptional and computational complexity of finite automata - a survey. *Inf. Comput.*, 209(3):456–470, 2011.