

Constructive Formalization of Regular Languages

Jan-Oliver Kaiser

September 5, 2012

Contents

1	Introduction	4
1.1	Recent work	4
1	Coq and SSReflect	2
1.1	Coq	2
1.2	SSREFLECT	2
1.2.1	Finite Types and Ordinals	2
1.2.2	Boolean Reflection	2
1.2.3	Boolean Predicates	2
3	Decidable Languages	7
3.1	Definition	7
3.1.1	Operation on languages	7
3.2	Regular Languages	9
3.2.1	Regular Expressions	9
3.2.2	Deciding Language Membership	10
4	Finite Automata	12
4.1	Definition	12
4.1.1	Determinism and Non-Determinism	12
4.2	Connected Components	14
4.3	Emptiness	15
4.4	Deciding Equivalence of Finite Automata	15
4.5	Regular Expressions and Finite Automata	16
4.5.1	Regular Expressions to Finite Automata	16
4.5.2	Deciding Equivalence of Regular Expressions	16
4.5.3	Finite Automata to Regular Expressions	16
5	Myhill-Nerode	20
5.1	Definition	20
5.2	Finite Partitionings and Equivalence Classes	21
5.3	Minimizing Equivalence Classes	22
5.4	Finite Automata and Myhill-Nerode	23
5.4.1	Finite Automata to Myhill-Nerode	23

<i>CONTENTS</i>	3
5.4.2 Myhill-Nerode to Finite Automata	23
2 Conclusion	4
3 References	5

Chapter 2

Coqand SSReflect

We decided to employ the Small Scale Reflection Extension (**SSReflect**) for the **Coq** proof assistant. The most important factors in this decision were SSREFLECT’s excellent support for finite types, list operations and graphs. SSREFLECT also introduces an alternative scripting language that can often be used to shorten the bookkeeping overhead of proofs considerably.

2.1 Coq

2.2 SSReflect

2.2.1 Finite Types and Ordinals

The most important feature of SSREFLECT for our purpose are finite types. SSREFLECT provides boolean versions of the universal and existential quantifiers on finite types, **forallb** and **existsb**. We can compute the number of elements in a finite type F with $\#|F|$. **enum** gives a list of all items of a finite type. Finite types also come with enumeration functions which provide a consistent ordering. The corresponding functions are **enum.rank** and **enum.val**. The input of **enum.val** and the result of **enum.rank** are ordinals, i.e. values in $[0, \#|F|-1]$. The corresponding type can be written as $1_{\#|F|}$.

2.2.2 Boolean Reflection

SSREFLECT offers boolean reflections for decidable propositions. This allows us to switch back and forth between equivalent boolean and propositional predicates.

2.2.3 Boolean Predicates

We make use of SSREFLECT’s syntax to specify boolean predicates. This allows us to specify predicates in a way that resembles set-theoretic nota-

tion, e.g. `[pred x | <boolean expression in x>]`. The complement of a predicate can be written as `[predC p]`. The syntax for combining predicates is `[pred? p1 & p2]`, with `?` being replaced with one of `U` (union), `I` (intersection) or `D` (difference). The resulting predicates are called collective predicates. SSREFLECT allows us to use infix notation, i.e. $x \in p$, for these predicates.

There are also applicative (functional) versions of `predC`, `predU`, `predI`, `predD` which are functions that take predicates and return applicative predicates.

Chapter 2

Conclusion

Chapter 3

References