

Functional Automata

Formal Languages for Computer Science Students

Marco T. Morazán
Seton Hall University
morazanm@shu.edu

Rosario Antunez
City College of New York
mrantunez@gmail.com

An introductory formal languages course exposes advanced undergraduate and early graduate students to automata theory, grammars, constructive proofs, computability, and decidability. Programming students find these topics to be challenging or, in many cases, overwhelming and on the fringe of Computer Science. The existence of this perception is not completely absurd since students are asked to design and prove correct machines and grammars without being able to experiment nor get immediate feedback, which is essential in a learning context. This article puts forth the thesis that the theory of computation ought to be taught using tools for actually building computations. It describes the implementation and the classroom use of a library, FSM, designed to provide students with the opportunity to experiment and test their designs using state machines, grammars, and regular expressions. Students are able to perform random testing before proceeding with a formal proof of correctness. That is, students can test their designs much like they do in a programming course. In addition, the library easily allows students to implement the algorithms they develop as part of the constructive proofs they write. Providing students with this ability ought to be a new trend in the formal languages classroom.

1 Introduction

An introductory formal languages course exposes advanced undergraduate and early graduate students to automata theory, grammars, constructive proofs, computability, and decidability. Programming-oriented students find these topics to be challenging or, in many cases, overwhelming and on the fringe of Computer Science. The existence of this perception is not completely absurd since students are asked to design and prove correct machines/grammars without being able to experiment nor get immediate feedback, as they do in any programming course, which is essential in a learning context and needs to be timely (the sooner the better) [9]. Computer Science students are accustomed to immediate feedback as they routinely use compilers and interpreters that provide error messages and warnings. Moreover, machines/grammars are representations of programs. Thus, in essence, designing machines and grammars without being able to test them goes against the grain of what students have learned in programming courses. The same holds true when arguing that a problem is decidable. In our experience, the inability to implement state machines and grammars results in inexperienced students submitting solutions that are incorrect even when an exercise is relatively simple. Experimentation and immediate feedback are key components to learning that are seldom found in the automata theory classroom.

Rarely do formal languages textbooks (e.g., [5, 7, 11]), unlike programming textbooks, offer any software infrastructure for students to experiment with their ideas and designs. This omission has an impact on how well students understand constructive proofs. In essence, a constructive proof spells out an algorithm to build a machine/grammar. For example, algorithms are outlined by the proof that nondeterministic finite-state machines and deterministic finite-state machines are equivalent and by the proof that context-free languages are closed under union. Computer Science students are taught that

algorithms need to be implemented and tested. This is especially true when the student has an intuition of how an algorithm ought to proceed, but is uncertain of all the details—the typical case when a student starts thinking about a constructive proof. Without feedback obtained from implementation and testing, it is common for students to try to prove the correctness of a machine/grammar that is incorrect. This leads to their work being marked down by an instructor and, in turn, to frustration and apathy for the material. The reader can contrast this scenario with a programming course in which the student can experiment with an implementation. Formal languages textbooks that reference software support (e.g., [6]) fail to integrate the use of the software as a part of the textbook. Instead, it is left as a recommendation. More worrisome, however, is that the support software is not easily extendible by students to integrate the algorithms they develop as part of their own constructive proofs. In other words, students are required to design algorithms that are not to be implemented.

Given that it is reasonable for Computer Science students to be able to experiment with and get feedback on the algorithms they develop, a decision must be made as to how to engage programming students. One solution, of course, is to have students implement the algorithms from scratch in their favorite programming language without providing any software support. Although possible, this approach is likely to be too time-consuming within the confines of one semester. A more reasonable approach is to provide students with a library (or a programming language) that allows them to quickly implement and test¹ the machines/grammars they design and the algorithms they develop. This article describes a Racket library, FSM, designed to provide students with the opportunity to implement and to test their designs using finite-state machines, regular expressions, regular grammars, pushdown automata, context-free grammars, Turing machines, and context-sensitive grammars much like they experiment when writing programs in a programming course. In addition, the FSM library easily allows students to implement the algorithms they develop as part of the constructive proofs they write. In essence, this article puts forth that the theory of computation ought to be taught using tools for actually building computations. Providing this ability should be a new trend in the formal languages classroom and in the development of software to support such courses. This software support allows students to receive immediate feedback on their work and, thus, the learning experience is enhanced. Furthermore, it is likely to provide significant help to instructors that commonly need to grade contrived machine designs for which determining correctness is difficult. This approach ties in nicely with other topics in a Computer Science curriculum, such as programming languages and compilers, by exposing students to programming with the data representation of programs.

The library provides the ability to replace pencil-and-paper designs with programming by making it an integral part of an Automata Theory course. Students use the library to implement and test all their designs and algorithms. The library has been implemented in Racket, because the resulting machines/grammars descriptions are concise. Furthermore, most of our students are familiar with the language syntax and with list-based processing. The reader can note that the library is also useful for students not familiar with Racket as only a minimal amount of syntax needs to be learned. Knowledge of functional programming is not a prerequisite to use FSM.

The article is organized as follows. Section 2 highlights the interface and the implementation of the FSM library to provide the framework needed by any reader that desires to implement the library using a different programming language. Section 3 provides examples of how the library has been used in practice and highlights the usefulness of immediate feedback in the formal languages classroom. Section 4 describes and contrast FSM with related work. Finally, Section 5 presents concluding remarks and directions for future work.

¹Tests provide a form of immediate feedback that either confirms or challenges the understanding of a student.

2 Interface and Implementation

The FSM library presents the user with a generic interface to construct and manipulate state machines and grammars. The interface is presented to facilitate its reproduction in any modern programming language. Constructors are divided into two categories: primitive constructors and transformers. Primitive constructors build a state machine or a grammar from a formal description provided by the programmer. Transformers build a state machine or a grammar from existing machines or grammars exploiting algorithms obtained from constructive proofs. Observers are divided into three categories: accessors, applicators, and testers. Accessors return a specified component used to build a grammar or a state machine. Applicators apply a given machine or grammar to a word. Testers allow for machines and grammars to be tested with words provided by the programmer or with randomly generated words by the software. The latter two provide students with immediate feedback on the validity (not the verification) of their designs and implementations.

2.1 State Machines

A state machine (sm) is either:

1. Deterministic finite-state machine (dfa)
2. Nondeterministic finite-state machine (ndfa)
3. Pushdown automata (pda)
4. Turing machine (tm)

Every state machine has the user explicitly provide a finite set of states (S), a tape alphabet (Σ), a starting state (s), a set of final states (F), and a set of transition rules (δ). The transition rules must describe a function for a dfa while simply a relation for the other machines. Finally, a pda also requires a stack alphabet (Γ). $|M|$ denotes the representation of a regular expression or a machine/grammar of type M . The corresponding primitive constructors for machines with explicit rules have the following signatures²:

1. **make-dfa**: $S \Sigma s F \delta_{dfa} \rightarrow |dfa|$,
 $(f b t) \in \delta_{dfa}$ where $f, t \in S$ and $b \in \Sigma$
2. **make-ndfa**: $S \Sigma s F \delta_{ndfa} \rightarrow |ndfa|$,
 $(f b t) \in \delta_{ndfa}$ where $f, t \in S$ and $b \in \{\Sigma \cup \{\varepsilon\}\}$
3. **make-pda**: $S \Sigma \Gamma s F \delta_{pda} \rightarrow |pda|$,
 $((f b g)(t l)) \in \delta_{pda}$ where $f, t \in S$, $b \in \{\Sigma \cup \{\varepsilon\}\}$, and
 $g, l \in \{\{\text{non-empty list of } \Gamma\text{-symbols}\} \cup \{\varepsilon\}\}$
4. **make-tm**: $S \Sigma \delta_{tm} s F \rightarrow |tm|$,
 $((f b)(t l)) \in \delta_{tm}$ where $f, t \in S$ and $b, l \in \{\Sigma \cup \{\varepsilon, \sqcup\}\}$

State machines are represented as functions that dispatch to the appropriate observer given some input.

FSM also employs a more general notion of a machine. A machine (m) is either:

1. A state machine
2. Combined Turing machine (ctm)³.

² ε represents the empty string and \sqcup represents a blank tape space.

³In terms of computational power there is no difference between a Turing Machine and a combined Turing machine. The latter is simply an abstraction that simplifies the design of Turing machines built from existing Turing Machines.

A *ctm* is not defined by providing a formal description. Instead, the user provides a *ctm* description (*ctmd*) and Σ . A *ctm* description uses other Turing machines as building blocks. In essence, a *ctm* represents an iterative algorithm with conditional branches and gotos that changes the state of the machine. A *ctmd* is described by the following grammar:

$$\begin{aligned}
 \langle ctmd \rangle &\rightarrow () \\
 &\rightarrow (\langle tm \rangle \langle ctmd \rangle) \\
 &\rightarrow (\langle ctmd \rangle \langle ctmd \rangle) \\
 &\rightarrow (\langle label \rangle \langle ctmd \rangle) \\
 &\rightarrow (\text{BRANCH } \{ \langle symbol \rangle \langle ctmd \rangle \}^+) \\
 &\rightarrow (\text{GOTO } \langle label \rangle) \\
 &\rightarrow ((\text{VAR } \langle symbol \rangle) \langle ctmd \rangle)
 \end{aligned}$$

Informally, the varieties of *ctmds* are described as follows. The simplest *ctmds* are either: empty, a *tm* followed by a *ctmd*, a *ctmd* followed by a *ctmd*, or a label (used by a goto) followed by a *ctmd*. More complex *ctmds*, can either start with a branch, a goto, or a variable. A branch defines options for the machine to execute based on the alphabet symbol being read. A goto unconditionally branches to a label that must be defined. Finally, to aid in the development of these machines, the library allows for branches to abstract over the symbol being read. A variable (not an alphabet symbol) captures the read symbol. This variable is used as a constant, whose value can be written to the tape, in the *ctm* that follows it⁴.

The constructor signature for a *ctm* is: **combine-tms**: $ctmd \Sigma \rightarrow ctm$. The observer **apply-ctm** takes as input a *ctm*, a tape configuration (akin to a word), and a head position and returns a Turing machine configuration (i.e., a state, a tape, and the position of the head on the tape). A *ctm* is represented as a function that returns a Turing machine configuration. In this manner, ctm_1 and ctm_2 can be composed as the resulting tape and head position of ctm_1 is used to build the initial configuration for ctm_2 .

The state-machine transformers build new machines from existing machines, from grammars, or from a regular expression. These constructors implement algorithms from constructive proofs typically covered in an introductory formal languages course. They include:

(regexp \rightarrow fsa | *regexp*) Transforms a regular expression into a finite-state machine.

(ndfa \rightarrow dfa | *ndfa*) Transforms a *ndfa* into a *dfa*.

(rename-states-sm (listof state) | *sm*) Renames the states of the given *sm* such that the intersection of the new names and the given list of states (i.e., symbols) is empty⁵. This function is useful when combining two state machines requires that the intersection of the set of states of both machines be empty like, for example, when creating a machine using closure under union.

(union-sm | *sm* | *sm*) Builds an *sm* that accepts the union of the languages of the two given *sms* of the same kind. If the inputs are *tms*, they must be language recognizers.

(concat-sm | *sm* | *sm*) Builds an *sm* that accepts the concatenation of the languages of the two given *sms* of the same kind. If the inputs are *tms*, they must be a language recognizers.

(kleenestar-sm | *sm*) Builds an *sm* that accepts the Kleene star of the given *sm*'s language. If the input is a *tm*, it must be a language recognizer.

(complement-sm | *sm*) Builds an *sm* that accepts the complement of the language of the given *sm*. The given *sm* cannot be a *pda*. If the input is a *tm*, it must be a language recognizer.

⁴In programming languages, this is akin to β -conversion.

⁵In programming languages, this is akin to α -conversion.

(intersection-sm $|sm| |sm|$) Builds an sm that accepts the intersection of the languages of the two given sms of the same kind. The given sms cannot be pdas. If the inputs are tms, they must be language recognizers.

(grammar→sm $|grammar|$) Builds an sm for the language of the given grammar.

The observers to return a given component of an sm take as input an sm and return the desired component. The interesting observers are:

(apply-sm $|sm| word \{natnum\}$) Runs the given sm assuming the given word is on the tape and the head of the sm is on position 0 of the tape. If the optional natural number is provided, the head starts at that position. The returned value is either 'accept or 'reject. The given sm is simulated by performing a breadth-first search of all the possible paths it can take by consuming the given word.

(show-transitions-sm $|sm| word \{natnum\}$) Similar to apply-sm, but returns the path followed by the sm. For a nondeterministic sm, it returns an empty path if all possible paths end in a non-accepting state.

The testers for state machines offer the user the ability to experiment with the machines they create and provide immediate feedback. There are three testers provided:

(same-result-sm? $|sm| |sm| word$) Determines if both of the given sms produce the same result for the given word.

(test-equiv-sm $|sm| |sm| \{natnum\}$) Determines if both of the given sms produce the same result on 100 randomly generated words. If the optional natural number is provided, then that number of random tests are performed. The function returns true or a list of words for which the given machines produce a different result. If a tm is provided as input, it must be a language recognizer.

(test-sm $|sm| \{natnum\}$) Generates 100 (or the given optional number of) random words and returns a list of the words with the results obtained from the given sm. If a tm is provided as input, it must be a language recognizer.

2.2 Grammars

A grammar is either:

1. Regular grammar (rg)
2. Context-free grammar (cfg)
3. Context-sensitive grammar (csg)

Every grammar is composed of a set of terminal and nonterminal symbols (V), a set of terminal symbols (Σ), a set of production rules (R), and a starting nonterminal symbol S . The left hand side of production rules for an rg and a cfg must only have a single nonterminal. The left hand side of production rules for a csg must contain at least one nonterminal. The right hand side of a production rule for a rg must have either a terminal symbol, a terminal symbol followed by a nonterminal symbol, or, if the left hand side is S , ϵ (the empty string). The right hand side of a production rule for a cfg and a csg may be ϵ or may have one or more members of V . The primitive constructors for grammars are:

1. **make-rg:** $V \Sigma R_{rg} S \rightarrow |rg|$,
 $(A \rightarrow aB) \vee (A \rightarrow a) \vee (S \rightarrow \epsilon) \in R_{rg}$ where $A, B \in V - \Sigma$ and $a \in \Sigma$
2. **make-cfg:** $V \Sigma R_{cfg} S \rightarrow |cfg|$,
 $(A \rightarrow a^+) \vee (A \rightarrow \epsilon) \in R_{cfg}$ where $A \in V - \Sigma$ and $a \in V$

3. **make-csg**: $V \Sigma R_{csg} \text{ symbol} \rightarrow |csg|$,
 $BTC \rightarrow (D \vee \varepsilon) \in R_{csg}$ where $B, C, D \in V^+$ and $T \in V - \Sigma$

The transformers for grammars build new grammars from either existing grammars, existing state machines, or an existing regular expression. They include:

- (sm→grammar |sm|)** Transforms a state machine into a grammar. If the input is a tm, it must be a language recognizer.
- (grammar-rename-nts (listofsymbol) |grammar|)** Renames the nonterminals of the given grammar to symbols not contained in the given list. This function is useful when the intersection of the nonterminals of two grammars must be empty.

The observers to return a given component of a grammar take as input a grammar and return the desired component. The interesting observer is:

- (deriv |grammar| word)** If the given word is in the language of the given grammar, the derivation of the word is returned. Otherwise, a string stating that the word is not a member of the language of the grammar is returned.

The testers for grammars offer students the ability to experiment with the grammars they create and offer immediate feedback. There are three testers provided:

- (both-deriv? |grammar| |grammar| word)** Determines if both of the given grammars derive the given word.
- (test-equiv-grammar |grammar| |grammar| {natnum})** Determines if both of the given grammars produce the same result when attempting to derive 100 randomly generated words. If the optional natural number is provided, then that number of random tests are performed. The function returns true or a list of words for which the given grammars produce a different result.
- (test-grammar |grammar| {natnum})** Generates 100 (or the given optional number of) random words and returns a list of the words with the results obtained from trying to use the given grammar to derive them.

2.3 Regular Expressions

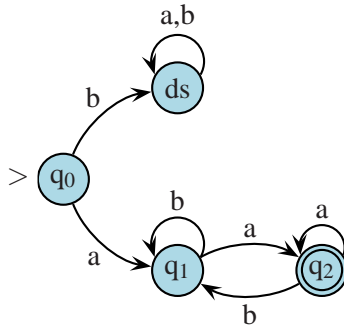
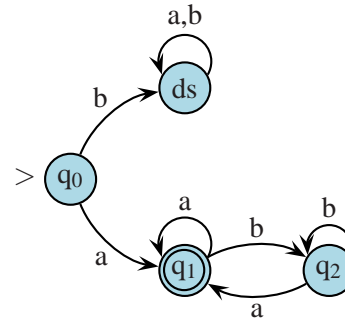
Let Σ be an alphabet of symbols. A regular expression (*re*) is a string that is either

1. ε (the empty string)
2. $a \in \Sigma$
3. $(re_1 \cup re_2)$, where re_1 and re_2 are *re*
4. $(re_1 re_2)$, where re_1 and re_2 are *re*
5. re^*

There is a constructor for each of the above. There is a single transformer, **(fsa→regexp |fsa|)**, that converts an ndfa to a regular expression. There is a single observer, **(printable-regexp r)**, that takes as input a regular expression and that returns a string representing the given regular expression.

3 The FSM Library in Practice

This section presents examples of the FSM library in practice using problems that students have faced in an introductory formal languages course. The examples highlight how students and instructors make the use of the library relevant to both and how immediate feedback enhances the learning experience.

Figure 1: Buggy Finite-State Automaton for L .Figure 2: Correct Finite-State Automaton for L .

3.1 Write a Program to Recognize a Regular Language

Let $\Sigma = \{a, b\}$. Consider the problem of recognizing the regular language:

$$L = \{w \mid w \in \Sigma^* \wedge w \text{ starts and ends with an } a\}.$$

It is not uncommon for a student to submit the automaton in Figure 1 which is clearly incorrect and earns the student a poor grade. The poor grade usually leads a student to experience frustration and apathy for the material. This frustration stems from realizing that the design has what they consider a small bug that experimentation and immediate feedback easily uncovers.

If the student has access to the FSM library, an implementation of their design looks as follows:

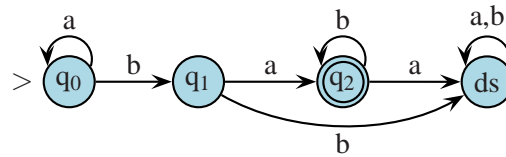
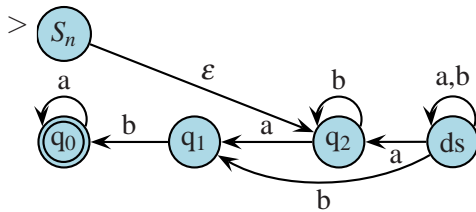
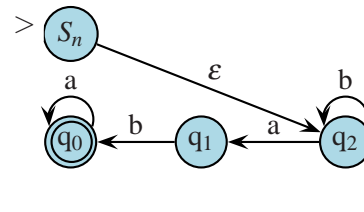
```

(define sol1-dfa (make-dfa '(q0 q1 q2 ds)
  '(a b)
  'q0
  '(q2)
  '((q0 a q1)
    (q0 b ds)
    (q1 a q2)
    (q1 b q1)
    (q2 a q2)
    (q2 b q1)
    (ds a ds)
    (ds b ds))))
  
```

The student can now test this machine as follows:

```

> (test-sm sol1-dfa)
'(((b b a a b b b b a b a b b b b a b) reject)
  ((b a b a a a b a b a a a a a b b b a b) reject)
  ((a a b a b a a b a a b b b b) reject)
  ((a b a a b b a b b b a) accept)
  ((a b b a) accept)
  ...
  ((a) reject)
  ...
  ((b a a a b b b b a a b b b a b) reject)
  ((b a b a a b b b) reject))
  
```

Figure 3: A dfa for $L = a^*bab^*$.Figure 4: M_{rev} from Theorem Version I.Figure 5: M_{rev} from Theorem Version II.

The tests reveal that the word '(a) is rejected, but this word is an element of L . The student can now refine the solution to the one displayed in Figure 2. The corresponding implementation is:

```
(define sol1-dfa (make-dfa '(q0 q1 q2 ds)
  '(a b)
  'q0
  '(q1)
  '((q0 a q1)
    (q0 b ds)
    (q1 a q1)
    (q1 b q2)
    (q2 a q1)
    (q2 b q2)
    (ds a ds)
    (ds b ds))))
```

This example may seem deceptively simple, but it does have a significant impact on a student's attitude. The immediate feedback provided by the FSM library assists students to find errors in their designs and to avoid the frustration of receiving a poor grade. Furthermore, the experience encourages both the student and the instructor resulting in an enhanced learning experience by eliminating the need to spend time on low-level bugs and allowing for more time to be dedicated to harder material.

3.2 The Reverse of a Regular Language

Proving that the reverse of a regular language, L_{rev} , is regular requires a constructive proof that students in an introductory formal languages course are commonly asked to write. Typically, students are confused about how to tackle this kind of problem. They understand that they must show how to build a finite-state automaton, M_{rev} , for L_{rev} , but feel frustrated if their algorithm is not correct. In a programming course, students can implement and experiment with their proposed solution. The same ought to be true for a course in formal languages and this is made possible by FSM.

A common proposed solution is to build M_{rev} from a deterministic finite-state automaton, M , for L . Intuitively, M_{rev} has M 's set of states and a new starting state, M 's alphabet, M 's transition rules reversed,

ε -transitions from the new starting state to each final state of M , and the starting state of M as its only final state. More formally students define M_{rev} in their proposed version I of the theorem as follows:

$M_{rev} = (\{Q_M \cup \{S_{rev}\}, \Sigma_M, \delta_{rev}, S_{rev}, \{S_M\}\})$, where

- Q_M = the states of M
- S_{rev} is the unique symbol for the starting state of M_{rev}
- Σ_M = the alphabet of M
- δ_{rev} contains two types of rules:
 1. $(S_{rev}, \varepsilon, q_i)$ such that $q_i \in F_M =$ the final states of M
 2. (q_i, σ, q_j) such that $(q_j, \sigma, q_i) \in \delta_M$
- S_M is the starting state of M

The above algorithm is implemented and tested by students using FSM. Consider applying the proposed theorem to a deterministic finite-state automaton for $L = a^*bab^*$ depicted in Figure 3. The resulting ndfa is depicted in Figure 4. Students can observe that the state ds is inaccessible and, therefore, ought not be part of the transformed machine. This observation suggests a refinement that eliminates the dead states of M and any transitions involving a dead state. Using this refinement, the resulting ndfa for M_{rev} is given in Figure 5. An FSM implementation of this second version of the proposed theorem is displayed in Figure 6. Once testing makes students confident that their solution is correct, they can proceed to write the formal proof.

Observe that in FSM new constructors are easily defined by students by simply writing a function. The code in Figure 6 only uses primitive functions on lists, λ -expressions, and list-based higher-order functions (e.g., map, filter, and andmap). Thus, the coding of this constructor ought to be well within the reach of both advanced undergraduate and beginning graduate Computer Science students. Students uncomfortable or unfamiliar with λ -expressions and higher-order functions can, instead, omit their use by explicitly defining functions for λ -expressions and by explicitly defining recursive functions to perform the necessary list-processing done by the higher-order functions.

3.3 Determining if a Context-Free Language is Empty

Students also face exercises to prove that a problem is decidable. If a problem is decidable, then there is an algorithm to determine if an instance of a given problem fulfills stipulated conditions. Students can propose an algorithm for decidability, but without the ability to test their solution they remain unsure about its validity. Testing their algorithm increases their confidence and brings their Computer Science education as programmers to bear in a formal languages course. We do not want students, however, to implement decidability algorithms using Turing machines. That would simply be too unwieldy. Instead, students ought to be able to use a library like FSM to implement and test their algorithms.

For example, consider the problem of deciding if the language, $L(G)$, of a context-free grammar, G , is empty. Students realize that $L(G)$ is not empty if there exists a derivation for any word formed by elements of the alphabet of G . Usually, they must be guided to realize that what they need is an algorithm to detect the existence of a derivation by creating any backward derivation to G 's starting symbol starting from the elements of the alphabet of G and ε . The algorithm accumulates Σ and the left hand sides of rules (i.e., nonterminal symbols) whose right hand sides only contain symbols in the accumulator. If at any step, G 's starting symbol is in the accumulator then $L(G)$ is not empty. If at any step, there are no new symbols to add to the accumulator then $L(G)$ is empty. Otherwise, the new symbols are added to the accumulator and the process recursively proceeds.

```

; dfa --> ndfa
(define (reverse-dfa m1)

;symbol (listof rules)--> boolean
(define (deadstate? s rules)
  (let ((fromrules (filter (lambda (r) (eq? s (smrule-fromstate r)))
                           rules)))
    (andmap (lambda (r) (eq? s (smrule-tostate r))) fromrules)))

(let* ((newfinal (sm-getstart m1))
      (mstates (sm-getstates m1))
      (newStart (gen-symbol 'S mstates))
      (deadsts (remove-duplicates
                (filter (lambda (a) (deadstate? a (sm-getdeltas m1)))
                       mstates)))
      (newQ (cons newStart (filter (lambda (q) (not (member q deadsts)))
                                   mstates)))
      (addedrules (map (lambda (s) (list newStart EMP s))
                      (sm-getfinals m1)))
      (changedrules
       (map reverse
            (filter (lambda (r) (and (not (member (smrule-tostate r)
                                                  deadsts))
                                   (not (member (smrule-fromstate r)
                                               deadsts)))))
                  (sm-getdeltas m1)))
      (newrules (append addedrules changedrules)))
  (make-ndfa newQ (sm-getsigma m1) newStart (list newfinal) newrules)))

```

Figure 6: Proposed solution to build M_{rev}

Figure 7 displays the implementation of this algorithm using the FSM library. Notice, that this code demonstrates that observers are easily added by simply writing a function. This observer only utilizes FSM provided functions and list-processing functions. Thus, once again, putting it well within the grasp of Computer Science graduate and advanced undergraduate students. Nonetheless, students tend to make mistakes at first and their algorithm requires refinement. The most common mistake is to not include ϵ in the initial value of the accumulator which is easily discovered once implemented.

3.4 Computing with Turing Machines

Although Turing machines are not the most attractive programming abstraction, it is important for students to understand their power and the reason the abstraction is less than attractive. The best way for students to begin to understand the power of Turing machines is to have them design Turing machines. Implementing formal descriptions of Turing machines in FSM is similar to developing the finite-state automata in Section 3.1. Such descriptions are best for language recognizers and operations that do

```

; cfg --> boolean
(define (Lcfg-isempty? g)
  ; cfg-rule (listof symbol) --> (listof cfg-rule)
  (define (only-accum-elems? rule accum)
    (and (not (member (cfg-rule-lhs rule) accum))
          (andmap (lambda (s) (member s accum)) (cfg-rule-rhs rule))))
  ; (listof cfg-rules) nonterminal (listof symbol) --> boolean
  (define (isempty? rls S accum)
    (cond [(member S accum) #f]
          [else
           (let ((newmembers (map cfg-rule-lhs
                                   (filter (lambda (r)
                                             (only-accum-elems? r accum))
                                           rls))))
             (cond [(empty? newmembers) #t]
                   [else (isempty? rls S (append newmembers accum))]))]))
  (let ((rls (cfg-get-the-rules g))
        (S (cfg-get-start g))
        (sigma (cfg-get-alphabet g)))
    (isempty? rls S (cons EMP sigma))))

```

Figure 7: A Function to Determine if the Language of a CFG is Empty.

not involve assignment (i.e., altering of the tape). For instance, the following Turing machine moves the head to the right of the current position of the head.

```

(define RI (make-tm '(s h)
                   '(I add1 sub1)
                   (list
                    (list (list 's 'I) (list 'h RIGHT))
                    (list (list 's 'add1) (list 'h RIGHT))
                    (list (list 's 'sub1) (list 'h RIGHT))
                    (list (list 's BLANK) (list 'h RIGHT)))
                   's
                   '(h)))

```

Computations with Turing machines, however, may require assignment and, thus, care during their development. In such a setting, offering abstractions is critical to keep students engaged. Many textbooks on formal languages develop a notation that is graphical and more transparent. In essence, the notation connects Turing machines, not states, using conditional branches and gotos. This allows for progressively more complex Turing machines to be designed from simpler Turing machines.

Consider the student having to add or subtract 1 from a non-zero unary number. The first step is to state the precondition and the postcondition for the Turing Machine. For example, the algorithm can be designed assuming the machine starts in the following configuration: $(op \sqcup number \sqcup)$, where op is either $add1$ or $sub1$, \sqcup denotes a blank space, and the head is on the first blank after the number. The machine stops in the following configuration: $(op \sqcup number \sqcup)$, where $number$ is the result of the computation. Assume that in addition to RI above the following simpler machines are also defined:

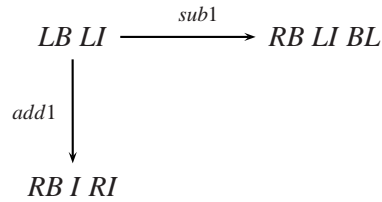


Figure 8: Draft Turing Machine to add or subtract 1 from a non-zero unary number.

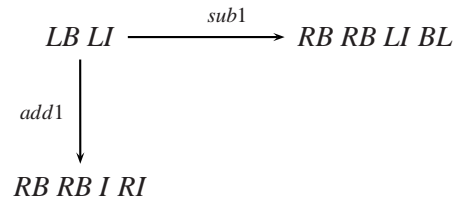


Figure 9: Turing Machine to add or subtract 1 from a non-zero unary number.

LI Moves the head one space to the left.

I Writes I to the tape.

BL Writes \square to the tape.

RB Moves the head to the first blank to the right of the head.

LB Moves the head to the first blank to the left of the head.

A first draft of the algorithm developed by a student is displayed in Figure 8. This algorithm moves the head to read the *op*. Then it branches depending on the *op*. After branching, it moves the head to the first blank to the right and proceeds to construct the resulting number. The FSM implementation is as follows:

```
(define addorsub (combine-tms (list LB LI (list BRANCH (list 'sub1 RB LI BL)
                                                         (list 'add1 RB I RI)))
                              '(I sub1 add1)))
```

This machine can be tested using `apply-sm` to obtain the following results:

```
> (apply-ctm addorsub (list add1 _IIII _) 6)
(tmconfig 'h 2 '(add1 IIIII _))
> (apply-ctm addorsub (list sub1 _IIIIIIII _) 9)
(tmconfig 'h 0 '(_ _IIII _))
```

A student quickly realizes that the postcondition of the machine is not met and, therefore, their design has a bug. Instead of handing in a buggy design and being marked down by the instructor, the student can now proceed to redesign their algorithm. The bug, of course, is that after the branch the machine must move to the second blank to the right. The resulting graphical notation is displayed in Figure 9. The FSM implementation is as follows:

```
(define addorsub (combine-tms (list LB LI (list BRANCH (list 'sub1 RB RB LI BL)
                                                         (list 'add1 RB RB I RI)))
                              '(I sub1 add1)))
```

The tests are repeated and the student gets the following results:

```
> (apply-sm addorsub (list add1 _IIII _) 6)
(tmconfig 'h 7 '(add1 _IIIII _))
> (apply-sm addorsub (list sub1 _IIIIIII _) 9)
(tmconfig 'h 8 '(sub1 _IIIIIII _))
```

With successful tests, the student can now proceed to develop arguments for correctness. This development establishes that the feedback provided by testing Turing machines is an important part of the design process in a formal languages course.

4 Related Work

JFLAP [10] was designed to experiment with state machines and grammars as well as to experiment with constructive proofs. JFLAP allows the user to create and simulate several types of state machines, to create and parse strings in the language of a grammar, and to experiment with proof constructions such as converting a nondeterministic finite automaton to a deterministic finite automaton and then to a regular expression or regular grammar [10]. A study concluded that students felt more engaged and enjoyed a formal languages course more when using JFLAP [10]. JFLAP provides all the primitive and transformation constructors found in FSM. In contrast, however, the graphical nature of JFLAP does not offer the ability to easily add new observers and constructors to the software neither does JFLAP generate random testing.

The jFAST [12] library assists beginners to design state machines. It uses a graphical interface just as JFLAP, but unlike JFLAP it provides no functionality for regular expressions and grammars. Like JFLAP and, in contrast to FSM, there is no support for students to add the constructive algorithms they develop and prove. The FSA Simulator allows the user to work and experiment with finite-state automata offering the ability to compare the languages of two finite-state automata [3]. That is, it provides testing facilities for the equivalence of two finite-state automata as FSM. The FSA comparison feature lets the software give students feedback about the accuracy of their work much as intended by FSM for all types of state machines. RegeXeX [1] is an interactive system to write regular expressions. In contrast to FSM, it provides testing facilities for regular expressions. The feedback provided by RegeXeX includes strings that ought to be accepted and ought to be rejected much like FSM provides testing facilities for state machines.

5 Conclusions and Future Work

The FSM library provides users the necessary facilities to design and experiment with state machines, grammars, and regular expressions. It supports the view that the existence of a machine or grammar is proven by developing a constructive proof. This means that the proof presents an algorithm that can be implemented using FSM. Students and instructors no longer have to rely solely on paper-and-pencil traces to build confidence or discover bugs in a design. Instead, they can use FSM testing facilities to generate tests that provide immediate feedback. This leads students to actively reason and learn about formal languages as well as to reinforce their Computer Science education by implementing and developing unit tests for their constructive algorithms. The library has received positive feedback by students and has provided the examples presented in the article. It is our expectation that the proposed approach, namely teaching the theory of computation with tools to build computations, be widely adopted by Computer Science programs.

Future work includes expanding the library to include more constructors particularly those for state minimization. We will also extend the library to include a graphical interface. Unlike the interfaces described in the related work, we do not wish to have students create machines and grammars using a graphical interface. Instead, our goal is to have students continue to write code to create machines and grammars that are then rendered using graphics to animate execution and visualize their structure. Additionally, more support for regular expressions and extensions of Turing machines will be offered. The latter machines, although not computationally more powerful than a standard Turing Machine, are likely to make certain designs easier to implement by students. Finally, a goal is to convert the library into a pedagogy-friendly embedded DSL (e.g., using Racket’s hygienic macros [4]) where it is possible to machine check the proofs (e.g., using tools like DrACuLa [2] and Coq [8]).

References

- [1] Christopher W. Brown & Eric A. Hardisty (2007): *RegeXeX: an interactive system providing regular expression exercises*. In: *SIGCSE '07: Proceedings of the 38th SIGCSE technical symposium on Computer science education*, ACM Press, New York, NY, USA, pp. 445–449, doi:10.1145/1227310.1227462.
- [2] Carl Eastlund & Matthias Felleisen (2009): *Automatic Verification for Interactive Graphical Programs*. In: *Proceedings of the Eighth International Workshop on the ACL2 Theorem Prover and Its Applications, ACL2 '09*, ACM, New York, NY, USA, pp. 33–41, doi:10.1145/1637837.1637843.
- [3] Michael T. Grinder (2003): *A Preliminary Empirical Evaluation of the Effectiveness of a Finite State Automaton Animator*. *SIGCSE Bull.* 35(1), pp. 157–161, doi:10.1145/792548.611958.
- [4] Shriram Krishnamurthi (2006): *EDUCATIONAL PEARL: Automata via Macros*. *J. Funct. Program.* 16(3), pp. 253–267, doi:10.1017/S0956796805005733.
- [5] Harry R. Lewis & Christos H. Papadimitriou (1997): *Elements of the Theory of Computation*, 2nd edition. Prentice Hall PTR, Upper Saddle River, NJ, USA.
- [6] P. Linz (2012): *An Introduction to Formal Languages and Automata*, 5th edition. Jones & Bartlett Learning.
- [7] J.C. Martin (2003): *Introduction to Languages and the Theory of Computation*. McGraw-Hill Series in Computer Science, McGraw-Hill.
- [8] The Coq Development Team (2004): *The Coq ProofAssistance Reference Manual*. LogiCal Project. Available at <http://coq.inria.fr>. Version 8.0.
- [9] Phil Race (2001): *Using Feedback to Help Students Learn*. Available at http://www.heacademy.ac.uk/resources/detail/resource_database/id432_using_feedback.
- [10] Susan H. Rodger, Eric Wiebe, Kyung Min Lee, Chris Morgan, Kareem Omar & Jonathan Su (2009): *Increasing Engagement in Automata Theory with JFLAP*. *SIGCSE Bull.* 41(1), pp. 403–407, doi:10.1145/1539024.1509011.
- [11] Michael Sipser (2013): *Introduction to the Theory of Computation*, 3rd edition. Cengage Learning.
- [12] Timothy M. White & Thomas P. Way (2006): *jfast: A java finite automata simulator*. In: *In Thirty-seventh SIGCSE Technical Symposium on Computer Science Education*, pp. 384–388, doi:10.1145/1121341.1121460.