An introduction to the FIRE engine: a C++ toolkit for finite automata and regular expressions

Citation for published version (APA):

Document status and date:
Published: 01/01/1994

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
An introduction to the FIRE engine:
A C++ toolkit for Finite automata
and Regular Expressions

by

B.W. Watson

94/21
This is a series of notes of the Computing Science Section of the Department of Mathematics and Computing Science Eindhoven University of Technology. Since many of these notes are preliminary versions or may be published elsewhere, they have a limited distribution only and are not for review. Copies of these notes are available from the author.

Copies can be ordered from:
Mrs. M. Philips
Eindhoven University of Technology
Department of Mathematics and Computing Science
P.O. Box 513
5600 MB EINDHOVEN
The Netherlands
ISSN 0926-4515

All rights reserved
editors: prof.dr.M.Rem
         prof.dr.K.M.van Hee.
An introduction to the FIRE engine:
A C++ toolkit for Finite automata and Regular Expressions*

Bruce W. Watson
Faculty of Mathematics and Computing Science
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
The Netherlands
email: watson@win.tue.nl

August 15, 1994

Abstract
This paper is an introduction to the programmer's interface of version 1.2 of the FIRE engine. The FIRE engine is a C++ class library implementing finite automata and regular expression algorithms. The algorithms implemented in the toolkit are almost all of those presented in the taxonomies of finite automata algorithms [Wat93a, Wat93b]. None of the implementation details of the library are discussed — such design and implementation details are given in [Wat94].

The toolkit is unique in providing implementations of all of the known algorithms for constructing finite automata. The implementations, which were developed with efficiency in mind, are intended for use in production quality applications. No shell or graphical user-interface is provided, as the toolkit is intended for integration into applications. The implementations of the algorithms follow directly from the abstract algorithms appearing in [Wat93a, Wat93b]. As such, the toolkit also serves as an educational example of the implementation of abstract algorithms.

*Second printing, with corrections.
CONTENTS

Contents

1 Introduction .................................................. 2
1.1 Related toolkits ............................................. 2
1.2 Advantages and characteristics of the FIRE engine .......... 3
1.3 Future directions for the toolkit ............................ 4
1.4 Obtaining and compiling the toolkit ......................... 4
1.5 Reading this paper .......................................... 5

2 Character ranges .............................................. 5

3 Regular expressions ......................................... 7

4 The Σ-algebra ................................................... 8

5 Abstract and concrete finite automata ......................... 10

6 Automata constructions ....................................... 11

7 Minimizing DFAs ............................................... 14

References ................................................................ 14
1 Introduction

The FIRE engine is a C++ class library implementing finite automata and regular expression algorithms. The algorithms implemented in the toolkit are almost all of those presented in the taxonomies of finite automata algorithms [Wat93a, Wat93b]. This paper serves as an introduction to the interface of version 1.2 of the FIRE engine. None of the implementation details of the library are discussed — such design and implementation details are given in [Wat94].

The toolkit is a computing engine, providing classes and algorithms of a low enough level that they can be used in most applications requiring finite automata or regular expressions. Unlike most of the other toolkits [RW93, CH91, JPTW90], the FIRE engine does not have an interactive user-interface. The toolkit (which consists of approximately 9000 lines of C++ code) is implemented as an object-oriented class library, as opposed to a procedural library, in order to reuse code as much as possible, and to minimize the cost of maintaining the code. C++ was chosen as the implementation language because compilers for C++ are more widely available than for any other object-oriented language.

The implementation of the FIRE engine was driven by a number of aims. Briefly, these aims are:

• To implement the algorithms efficiently enough for use in production quality applications.

• To demonstrate that the abstract algorithms appearing in [Wat93a, Wat93b] could indeed be easily (and efficiently) implemented.

• To provide implementations of all of the known algorithms; this will facilitate performance comparisons between the algorithms, and provide students with a variety of algorithms for educational purposes.

• To implement only a class library as a computing engine, without a user-interface.

The choice of these aims becomes clear when the characteristics of the other existing toolkits are considered.

The following subsections briefly outline other finite automata toolkits, the advantages and characteristics of the FIRE engine, future directions for the toolkit, information on obtaining and compiling the toolkit, and an outline of the structure of this paper.

Acknowledgements: I would like to thank the following people for their assistance in the preparation of this paper: Pieter 't Hoen (for implementing two of the algorithms from [Wat93b]), and Kees Hemerik, Erik Poll, Nanette Saes, Richard Watson, and Gerard Zwaan (for providing style suggestions on this paper).

1.1 Related toolkits

There are several existing finite automata toolkits. They are:

• The AMORE system, as described in [JPTW90]. The AMORE package is an implementation of the semigroup approach to formal languages. It provides procedures for the manipulation of regular expressions, finite automata, and finite semigroups. The system supports a graphical user-interface on a variety of platforms, allowing the user to interactively and graphically manipulate the finite automata. The program is written (portably) in the C programming language, but it does not provide a programmer's interface. The system is intended to serve two purposes: to support research into language theory and to help explore the efficient implementation of algorithms solving language theoretic problems.

• The AUTOMATE system, as described in [CH91]. AUTOMATE is a package for the symbolic computation on finite automata, extended regular expressions (those with the intersection and complementation operators), and finite semigroups. The system provides a textual
user-interface through which regular expressions and finite automata can be manipulated. A single finite automata construction algorithm (a variant of Thompson's) and a single deterministic finite automata minimization algorithm is provided (Hopcroft's). The system is intended for use in teaching and language theory research. The (monolithic) program is written (portably) in the C programming language, but provides no function library interface for programmers.

- The GRAIL system, as described in [RW93]. GRAIL follows in the tradition of such toolkits as REGPACK [Lei77] and INR [Joh86], which were all developed at the University of Waterloo, Canada. It provides two interfaces:
  - A set of "filter" programs (in the tradition of UNIX). Each of the filters implements an elementary operation on finite automata or regular expressions. Such operations include conversions from regular expressions to finite automata, minimization of finite automata, etc. The filters can be combined as a UNIX "pipe" to create more complex operations; the use of pipes allows the user to examine the intermediate results of complex operations. This interface satisfies the first two (of three) aims of GRAIL [RW93]: to provide a vehicle for research into language theoretic algorithms, and to facilitate teaching of language theory.
  - A raw C++ class library provides a wide variety of language theoretic objects and algorithms for manipulating them. The class library is used directly in the implementation of the filter programs. This interface is intended to satisfy the third aim of GRAIL: an efficient system for use in application software.

The provision of the C++ class interface in GRAIL makes it the only toolkit with aims similar to those of the FIRE engine.

1.2 Advantages and characteristics of the FIRE engine

The advantages to using the FIRE engine, and the similarities and differences between the FIRE engine and the existing toolkits are:

- The FIRE engine does not provide a user interface\(^1\). Many of the other toolkits provide user interfaces for the symbolic manipulation of finite automata and regular expressions. Since the FIRE engine is strictly a computing engine, it can be used as the implementation beneath a symbolic computation application.

- The toolkit is implemented for efficiency. Unlike the other toolkits, which are implemented with educational aims, it is intended that the implementations in the FIRE engine are efficient enough that they can be used in production quality software.

- Despite the emphasis on efficiency in the FIRE engine, the toolkit still has educational value. The toolkit bridges the gap between the easily understood abstract algorithms appearing in [Wat93a, Wat93b] and practical implementation of such algorithms. The C++ implementations of the algorithms display a close resemblance to their abstract counterparts.

- Since the algorithms are direct implementations of those appearing in the taxonomies of abstract algorithms [Wat93a, Wat93b], and correctness proofs are provided for the abstract algorithms in the taxonomies, there is a high degree of confidence in the correctness of the code in the FIRE engine.

- Most of the toolkits implement only one of the known algorithms solving any particular problem. For example, AUTOMATE implements only one of the known finite automata construction algorithms. The FIRE engine provides implementations of almost all of the known algorithms for all of the problems considered in the taxonomies [Wat93a, Wat93b]. Implementing all of the known algorithms has several advantages:

\(^1\)A rudimentary user interface is included for demonstration purposes.
- the efficiency of the algorithms (on a given application) can be compared.
- the algorithms can be studied and compared by those interested in the inner workings of the algorithms;

- The use of $\Sigma$-algebras in [Wat93a] can provide great computational efficiency in practice. For example, from regular expressions $E_0$ and $E_1$ we can construct finite automata $M_0, M_1$ (accepting the languages denoted by $E_0, E_1$ respectively). Assume that we now require a finite automaton accepting the language denoted by $E_0 \cdot E_1$ (their concatenation). With some of the existing toolkits, the new finite automaton would be constructed from scratch. With the FIRE engine, a concatenation operator on finite automata is implemented (for two of the different varieties of finite automata), enabling us to compute $M_0 \cdot M_1$ (a finite automaton accepting the desired language). This type of reuse of intermediate results can be a great computational saving.

The complete $\Sigma$-algebra as presented in [Wat93a] is implemented in the FIRE engine.

- The class hierarchy in the FIRE engine reflects the derivation of the taxonomies in [Wat93a, Wat93b]. In the taxonomies, common parts of algorithm derivations are factored out and presented as one. In the FIRE engine, such common parts of derivations are implemented as base classes (thereby implementing the code reuse embodied in the factored derivation).

1.3 Future directions for the toolkit

A number of improvements to the FIRE engine will appear in future versions:

- Presently, the FIRE engine implements only acceptors. Transducers (as would be required for some types of pattern matching, lexical analysis, and communicating finite automata) will be implemented in a future version.

- The FIRE engine only implements $\Sigma$-algebras for three carrier sets (regular expressions, finite automata, and reduced finite automata). Future versions of the toolkit will include more $\Sigma$-algebras.

- A future version of the toolkit will include support for extended regular expressions, i.e. regular expressions containing intersection or complementation operators.

- Basic regular expressions and automata transition labels are represented by character ranges. A future version of the FIRE engine will permit basic regular expressions and transition labels to be built from more complex data-structures. For example, it will be possible to process a string (vector) of structures. (Version 2.0 of GRAIL includes a similar improvement.)

1.4 Obtaining and compiling the toolkit

The FIRE engine can be obtained by anonymous ftp from Internet node ftp.win.tue.nl (also known as 131.155.70.100) in directory /pub/techreports/pi/automata/toolkit/v1.1

The FIRE engine makes use of C++ language features that are defined in the "ARM" version of the language [SE90]. The toolkit has been successfully compiled with Watcom C/C++ 32 version 9.5. The interdependencies between classes is complex, and is fully outlined in [Wat94]. When using only part of the toolkit, instead of determining the dependencies (and thus the required files for compiling and linking), the simplest approach is to compile the entire toolkit, and use a "smart linker" to remove any excess code or data from the executable image.

A makefile compatible with Watcom Make (WMAKE) is included in the distribution.
1.5 Reading this paper

To make use of the FIRE engine (and to read this paper), a basic working knowledge of C++ is required. For introductory treatments of C++, see [Bud91, Bud94, Lip91, Str91]. In order to fully understand the FIRE engine (by reading [Wat94]), it is necessary to read the taxonomies [Wat93a, Wat93b], and to have an advanced knowledge of C++ techniques as presented in [Cop92, MeS92, Mur93, Str91].

All references to "the taxonomy" appearing in the C++ code are references to [Wat94]. The code given in this paper only includes the public parts. For the private parts, see [Wat94].

This paper is structured as follows:

1. Section 2 gives a description of character ranges. Objects of this type are used as basic regular expressions and as finite automata transition labels.

2. Section 3 outlines the class used to implement regular expressions. Because of the nature of Σ-algebras [Wat93a, Section 3], this class does not provide a direct means of constructing regular expressions; such a means is provided by the Reg template class, described in Section 4. The regular expression class simply implements the terms of the Σ-term algebra (see [Wat93a, Section 3]).

3. Section 4 presents the Reg template class, which is used to implement the Σ-algebra given in [Wat93a, Section 3]. The template defines the operators of the Σ-algebra (see [Wat93a, Section 3]).

4. Section 5 presents finite automata. An abstract finite automata class is described. All concrete automata classes inherit from this class, which is used to define a common client interface. Concrete (non-abstract) automata classes are also described. These include: finite automata, reduced finite automata, left-biased finite automata, right-biased finite automata, and deterministic finite automata.

5. Section 6 gives the function prototypes of the finite automata construction functions (those functions taking a regular expression, and returning some type of finite automata).

6. Section 7 gives the function prototypes of the deterministic finite automata minimization functions.

Examples of the use of the classes are spread throughout the paper.

2 Character ranges

In the FIRE engine, atomic regular expressions and finite automata transitions can be labeled by sets of characters. The sets of characters allowed are restricted to non-empty ranges of characters in the execution character set (which is usually ASCII or EBCDIC). Class CharRange is used to represent such ranges of characters, which are specified by their upper and lower (inclusive) bounds.

The only public member functions relevant to the client interface are the constructors and the comparison operators. The argumentless constructor yields a CharRange, which denotes the empty range, i.e. one that is invalid. This constructor should not be used by clients. A second constructor takes a char and yields the singleton character range (a character range whose upper and lower bounds are equal) containing the char. The third constructor takes a pair of chars, the upper and lower bounds of a character range, and yields the CharRange denoting the range (since the empty range is not permitted, the upper bound must not be less than the lower bound). Constructing and using a CharRange denoting the empty range will eventually be caught by an assertion within the FIRE engine.

/* (c) Copyright 1994 by Bruce Watson */
2 CHARACTER RANGES

/* Revision: 1.2 */
/* Date: 1994/08/15 14:00:20 */
#ifndef CHARRANGE_H
#define CHARRANGE_H

#include "extras.h"
#include <cassert.h>
#include <iostream.h>

// Represent a set of characters (in the alphabet) in a subrange of char.
// Used by most classes as the basis of regular expressions and automata.

class CharRange {

public:

// Some constructors:
// The default is the empty range of characters.
// and this does not satisfy the class invariant.
inline CharRange();

// Copy constructor is standard.
inline CharRange( const CharRange& r );
inline CharRange( const char l, const char h );
inline CharRange( const char a );

// Default operator=, destructor are okay

// Some normal member functions:
// Is char a in the range?
inline int contains( const char a ) const;

// Access to the representation:
inline char low() const;
inline char high() const;

// Standard comparison operators.
inline int operator==( const CharRange& r ) const;
inline int operator!=( const CharRange& r ) const;

// Containment operators.
inline int operator<=( const CharRange& r ) const;
inline int operator>=( const CharRange& r ) const;

// Is there something in common between *this and r?
inline int not_disjoint( const CharRange& r ) const;

// Do *this and r overlap, or are they adjacent?
inline int overlap_or_adjacent( const CharRange& r ) const;

// Some operators on CharRanges.
// Merge two CharRanges if they're adjacent:
inline CharRange& merge( const CharRange& r );

// Make *this the intersection of *this and r.
inline CharRange& intersection( const CharRange& r );

// Return the left and right excess of *this with r (respectively):
inline CharRange left_excess( const CharRange& r ) const;
inline CharRange right_excess( const CharRange& r ) const;

// Define an ordering on CharRange's, used mainly in RE::ordering().
int ordering( const CharRange& r ) const;

// The class structural invariant.
inline int class_invariant() const;
};
3 Regular expressions

Class \texttt{RE} is used to represent regular expressions. This class cannot be used to construct regular expressions; to construct them, see Section 4. Class \texttt{RE} supports an assignment operator, comparison operators, and operators providing information on the structure of a regular expression (such as the main, or root operator of the regular expression). Most of the operators are used internally by other classes and functions of the \texttt{FIRE} engine. The argumentless constructor yields the regular expression denoting the empty language. Class \texttt{RE} also has input and output (\texttt{iostream} insertion and extraction) operators.

```cpp
// Some extra stuff.
friend ostream& operator<<( ostream& os, const CharRange r );
#endif

3 Regular expressions

class \texttt{RE} {
    public:
        // Some constructors, destructors, and operator=:
        // By default, create the RE denoting the empty language.
        \texttt{RE};
        \texttt{RE} ( const \texttt{RE}& r );
        \texttt{virtual } ^{\texttt{RE}}();
        \texttt{const } \texttt{RE}& \texttt{operator=}( \texttt{const } \texttt{RE}& \texttt{r} );
        // Some basic access member functions:
        // How many \texttt{SYMBOL} nodes in this regular expression?
        // Mainly for use in constructing \texttt{RFA}'s.
        \texttt{int } \texttt{num\_symbols()} \texttt{const};
        // How many operators in this RE?
        // Used in \texttt{ISImpl} (the items set stuff).
        \texttt{int } \texttt{num\_operators()} \texttt{const};
        // What is the main operator of this regular expression?
        \texttt{inline } \texttt{REops} \texttt{root\_operator()} \texttt{const};
        // What is the \texttt{CharRange} of this RE, if it is a \texttt{SYMBOL} regular expression.
        \texttt{inline } \texttt{CharRange} \texttt{symbol()} \texttt{const};
```
// What is the left RE of this RE operator (if it is a higher operator).
inline const RE& left_subexpr() const;

// What is the right RE of this RE operator (if it is a binary operator).
inline const RE& right_subexpr() const;

// Some derivatives (Brezouski's) related member functions:
// Does *this accept epsilon?
// This is from Definition 3.20 and Property 3.21
int Null() const;

// What is the First of *this?
// This is from Definition 4.60
CRSet First() const;

// Is *this in similarity normal form (SNF)?
int in_snf() const;

// Put *this into similarity normal form.
RE& snf();

// Reduce (optimize) *this by removing useless information.
RE& reduce();

// What is the derivative of *this (w.r.t. r)?
RE derivative( const CharRange& r ) const;

// Some ordering members (largely used in similarity and comparisons)
int ordering( const RE& r ) const;

// Some comparisons:
inline int operator==( const RE& r ) const;
inline int operator!=( const RE& r ) const;
inline int operator<( const RE& r ) const;
inline int operator<=( const RE& r ) const;
inline int operator>( const RE& r ) const;
inline int operator>=( const RE& r ) const;

// Some extras:
friend ostream& operator<<( ostream& os, const RE& r );
friend istream& operator>>( istream& is, RE& r );
inline int class_invariant() const;
);
#endif

4 The Σ-algebra

Template class Reg is used to define a client interface for the regular Σ-algebra (which is fully defined in [Wat93a, Section 3]). The template takes a single type parameter T which is used as the carrier set of the Σ-algebra. Class Reg<T> derives publicly from T, and so an Reg<T> can be used anywhere that a T can. The following member functions are provided:

- An argumentless constructor, a copy constructor, and an assignment operator (assuming that T has all three).
• A constructor from \( RE \) (a regular expression). Given that regular expressions are the \( \Sigma \)-term algebra, this constructor constructs the homomorphic image of the regular expression in the \( \Sigma \)-algebra of \( T \).

• \texttt{epsilon} makes \texttt{*this} accept the empty word only.

• \texttt{empty} makes \texttt{*this} accept the empty language.

• \texttt{symbol} takes a \texttt{CharRange} and makes \texttt{*this} accept the set of \texttt{chars} denoted by the \texttt{CharRange}.

• \texttt{or} takes another \texttt{Reg<T>}, \( r \), and makes \texttt{*this} accept the language of \texttt{*this} union the language of \( r \).

• \texttt{concat} takes another \texttt{Reg<T>}, \( r \), and makes \texttt{*this} accept the language of \texttt{*this} concatenated (on the right) with the language of \( r \).

• \texttt{star} makes \texttt{*this} accept the Kleene closure of the language of \texttt{*this}.

• \texttt{plus} makes \texttt{*this} accept the plus closure of the language of \texttt{*this}.

• \texttt{question} adds the empty word, \( e \), to the language accepted by \texttt{*this}.

For any two carrier sets \( T_1 \) and \( T_2 \), the \( \Sigma \)-algebra operator definitions will have very little in common. For this reason, the operators of template class \texttt{Reg} are specially defined defined for each carrier set. Presently, the operators are only defined for three type parameters (carrier sets): \texttt{Reg<RE>}, \texttt{Reg<FA>}, and \texttt{Reg<RFA>}.

```cpp
/* (c) Copyright 1994 by Bruce Watson */
// $Revision: 1.8 $
// $Date: 1994/08/15 14:00:50 $
ifndef SIGMA_H
#define SIGMA_H
#include "charrang.h"
#include "reaps.h"
#include "re.b"

// The signature of the Sigma-algebra is defined as template Reg. It is only
// an interface template, with most of the member functions being filled in as
// specialized versions. Given class T, the class Reg<T> is a Sigma-algebra with
// T as the carrier set. Reg is the only template required, since the regular
// expression Sigma algebra only has one sort: Reg.
// The Reg of a class T is also publicly derived from T, meaning that once
// a Reg<T> is constructed (perhaps using the Sigma-algebra operators), it can
// be used as a regular T (without slicing).
// A special constructor is provided, which constructs the homomorphic image in
// algebra Reg<T> of some regular expression. (Regular expressions are the Sigma-
// term algebra.)

template<class T>
class Reg : public T {
public:
   // Some constructors. Usually pass control back to the base class.
   Reg() : T() {}
   Reg( const Reg<T>& r ) :
      // Just pass back to the base class.
      T( (const T&)r ) {}
   // Construct the homomorphic image of regular expression r, using the Sigma-
   // algebra operators (corresponding to the operators of the regular expression).
   Reg( const RE& r ) { assert( r.class_invariant() );
```
homomorphic_image( r );
assert( T::class_invariant() );

// Default destructor falls through to the "T()"
inline const Reg<T>& operator=( const Reg<T>& r ) {
    // Call through to the base-class's operator=
    T::operator=( r );
    return( *this );
}

// Now the Sigma-algebra signature:

// Sigma-algebra basis
Reg<T>& epsilon();
Reg<T>& empty();
Reg<T>& symbol( const CharRange r );

// Sigma-algebra binary ops
Reg<T>& or( const Reg<T>& r );
Reg<T>& concat( const Reg<T>& r );

// Sigma-algebra unary ops
Reg<T>& star();
Reg<T>& plus();
Reg<T>& question();

#endif

5 Abstract and concrete finite automata

All finite automata in the FIRE engine share the same interface. This is enforced by deriving all concrete types of finite automata from the abstract base class FAabs, which defines the client interface. The interface includes member functions to restart the automaton (in its start states), advance the automaton by one char in the input string, determine if the automaton is in an accepting state, determine if the automaton is stuck (unable to make further transitions), and to determine if a particular string is accepted by the automaton (starting from the current state).

/* (c) Copyright 1994 by Bruce Watson */
// $Revision: 1.2 $
// $Date: 1994/08/15 14:00:42 $
 ifndef FAABS_H
 #define FAABS_H

 // Just mention DFA here, since circularity would be a problem if we include dfa.h. class DFA;

 // Abstract finite automaton. All other automaton types are derived from this one.
 // Basic FA operations are supported.

class FAabs {
public:
    // Return the number of states (or some reasonably close measure).
    virtual int num_states() const = 0;

    // Reset the current state before beginning to process a string.
    // This is not the default condition for most of the derived classes.
    virtual void restart() = 0;

    // Advance the current state by processing a character.
virtual void advance( char a ) = 0;

// Is the current state an accepting (final) one?
virtual int in_final() const = 0;

// Is the automaton stuck?
virtual int stuck() = 0;

// Is the string w acceptable?
virtual int acceptable( const char *w ) {
    for( restart(); !stuck() && *w; advance( *(w++) ) ) {}
    // It's acceptable if this is final, and the whole of w was consumed.
    return( in_final() && !*w );
}

// Return a DFA accepting the same language as *this.
virtual DFA determinism() const = 0;

#endif


Five concrete finite automata classes are provided in the FIRE engine: FA, RFA, LBFA, RBFA, and DFA (their interfaces are defined in files fa.h, rfa.h, lbfa.h, rbfa.h, and dfa.h respectively). The classes (respectively) represent general, reduced, left-biased, right-biased, and deterministic finite automata. They are all concrete because they inherit from abstract class FAabs, and do not leave any of the member functions as pure virtual member functions (each of the classes implements the client interface defined in class FAabs).

Although some of the classes provide some extra member functions, most of these members are for internal use by the FIRE engine. A full explanation of these member functions is given in [Wat94]. The construction of any of the concrete finite automata is easily done with the construction functions described in Section 6.

6 Automata constructions

Three header files contain (inline) definitions or prototypes of functions that take a regular expression and construct a finite automaton accepting the language of the regular expression. These functions implement all of the construction algorithms given in [Wat93a]. They differ in the type of finite automaton that is created:

- File constrs.h contains those construction functions whose operation is based upon the structure of a regular expression (the Σ-algebra functions). These functions correspond to those given in [Wat93a, Section 4].
- File dconstrs.h contains functions implementing Brzozowski's construction (and its variants). These functions all produce a DFA, and correspond to those given in [Wat93a, Section 5.3].
- File iconstrs.h contains the item set construction (and its variants). These functions all produce a DFA, and correspond to those given in [Wat93a, Section 5.5].
#include "lbfa.h"
#include "rbfa.h"
#include "sigma.h"
#include "dfa.h"
#include <assert.h>

// Thompson's construction:
// use the functions Thompson or Thompson_sigma, or
// auto FA Th( E );
// auto Reg<FA> Th( E );

// Construction 4.5
inline FA Thompson( const RE& r ) {
    return( FA( r ) );
}

// Construction 4.3
inline FA Thompson_sigma( const RE& r ) {
    return( Reg<FA>( r ) );
}

// RFA constructions:
// use the functions, or
// auto RFA R( E );
// auto Reg<RFA> R( E );
// To DFA:
// R.determinism();
// R2.determinism();
// R.determinism2();
// R2.determinism2();

// Definition 4.30 (variation)
inline RFA rfa( const RE& r ) {
    return( RFA( r ) );
}

// Definition 4.30
inline RFA rfa_sigma( const RE& r ) {
    return( Reg<RFA>( r ) );
}

// Construction 4.39 (variation of McNaughton-Yamada-Glushkov)
inline DFA MYG_shortcut( const RE& r ) {
    return( RFA( r ).determinism20 );
}

// Construction 4.50 (variation of Aho-Sethi-Ullman)
inline DFA ASU_shortcut( const RE& r ) {
    return( RFA( r ).determinism() );
}

// LBFA constructions:
// using RFA's.

// Construction 4.32
inline LBFA BerrySethi( const RE& r ) {
    return( RFA( r ) );
}

// Construction 4.38
LBFA BS_variation( const RE& r );

// DFA from LBFA construction:
// Construction 4.39
inline DFA MYG( const RE& r ) {
    return( BerrySethi( r ).determinism() );
}
6 AUTOMATA CONSTRUCTIONS

```c
// RBFA constructions:
// using RFAs'.

// Construction 4.45
inline RBFA BerrySethi_dual( const RE& r ) {
    return( RFA( r ) );
}

// Construction 4.48
RBFA BS_variation_dual( const RE& r );

// Construction 4.50
inline DFA ASU( const RE& r ) {
    return( BS_variation_dual( r ).determinism() );
}
#endif

/* (c) Copyright 1994 by Bruce W. Watson */
/* $Revision: 1.2 $ */
/* $Date: 1994/08/15 14:00:26 $ */
#ifndef DCONSTRS_H
#define DCONSTRS_H
#include <re.h>
#include "dfa.h"
#include <assert.h>
#endif

// The numbers in parentheses refer to the Construction number in the Taxonomy.

// DFA constructions:
// Brzozowski's constructions:
// Normal (Construction 5.34)
DFA Brz( const RE& r );

// With strong similarity (Construction 5.34 + Remark 5.32)
DFA Brz_optimized( const RE& r );
#endif

/* (c) Copyright 1994 by Bruce W. Watson */
/* $Revision: 1.2 $ */
/* $Date: 1994/08/15 14:00:43 $ */
#ifndef ICONSTRS_H
#define ICONSTRS_H
#include "re.h"
#include "dfa.h"
#include <assert.h>
#endif

// The numbers in parentheses refer to the Construction number in the Taxonomy.

// Item set constructions:
// Iconstr (Construction 5.69)
DFA Iconstr( const RE* r );

// DeRemer's (Construction 5.75)
DFA DeRemer( const RE* r );

// Oconstr (Construction 5.82)
DFA Oconstr( const RE* r );
```
7 Minimizing DFAs

Five deterministic finite automata minimization algorithms are implemented in the toolkit. All of the algorithms are implemented as member functions of class DFA, and are therefore declared in file dfa.h. The algorithms are:

- Brzozowski's, given in [Wat93b, Section 2]. It appears as member function min_Brzozowski.
- Aho, Sethi, and Ullman’s, given in [Wat93b, Algorithm 4.6]. It appears as member function min_dragon.
- Hopcroft and Ullman’s, given in [Wat93b, Algorithm 4.7]. It appears as member function min_HopcroftUllman.
- Hopcroft’s, given in [Wat93b, Algorithm 4.8]. It appears as member function min_Hopcroft.
- A new algorithm, given in [Wat93b, Algorithm 4.9, 4.10]. It appears as member function min_Watson.

References


REFERENCES


<table>
<thead>
<tr>
<th>No.</th>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/02</td>
<td>R.P. Nederpelt, H.C.M. de Swart</td>
<td>Implication. A survey of the different logical analyses &quot;if...then...&quot;, p. 26.</td>
</tr>
<tr>
<td>91/03</td>
<td>J.P. Katoen, L.A.M. Schoenmakers</td>
<td>Parallel Programs for the Recognition of P-invariant Segments, p. 16.</td>
</tr>
<tr>
<td>91/05</td>
<td>D. de Reus</td>
<td>An Implementation Model for GOOD, p. 18.</td>
</tr>
<tr>
<td>91/06</td>
<td>K.M. van Hee</td>
<td>SPECIFICATIEMETHODEN, een overzicht, p. 20.</td>
</tr>
<tr>
<td>91/07</td>
<td>E. Poll</td>
<td>CPO-models for second order lambda calculus with recursive types and subtyping, p. 49.</td>
</tr>
<tr>
<td>91/12</td>
<td>E. van der Sluis</td>
<td>A parallel local search algorithm for the travelling salesman problem, p. 12.</td>
</tr>
<tr>
<td>91/14</td>
<td>P. Lemmens</td>
<td>The PDB Hypermedia Package. Why and how it was built, p. 63.</td>
</tr>
<tr>
<td>91/16</td>
<td>A.J.J.M. Marcelis</td>
<td>An example of proving attribute grammars correct: the representation of arithmetical expressions by DAGs, p. 25.</td>
</tr>
<tr>
<td>Page</td>
<td>Authors</td>
<td>Title</td>
</tr>
<tr>
<td>------</td>
<td>-----------------------</td>
<td>-----------------------------------------------------------------------</td>
</tr>
<tr>
<td>91/18</td>
<td>Rik van Geldrop</td>
<td>Transformational Query Solving, p. 35.</td>
</tr>
<tr>
<td>91/19</td>
<td>Erik Poll</td>
<td>Some categorical properties for a model for second order lambda calculus with subtyping, p. 21.</td>
</tr>
<tr>
<td>91/23</td>
<td>K.M. van Hee, L.J. Somers, M. Voorhoeve</td>
<td>Z and high level Petri nets, p. 16.</td>
</tr>
<tr>
<td>91/24</td>
<td>A.T.M. Aerts, D. de Reus</td>
<td>Formal semantics for BRM with examples, p. 25.</td>
</tr>
<tr>
<td>91/25</td>
<td>P. Zhou, J. Hooman, R. Kuiper</td>
<td>A compositional proof system for real-time systems based on explicit clock temporal logic: soundness and completeness, p. 52.</td>
</tr>
<tr>
<td>91/27</td>
<td>F. de Boer, C. Palamidessi</td>
<td>Embedding as a tool for language comparison: On the CSP hierarchy, p. 17.</td>
</tr>
<tr>
<td>91/28</td>
<td>F. de Boer</td>
<td>A compositional proof system for dynamic process creation, p. 24.</td>
</tr>
<tr>
<td>91/30</td>
<td>J.C.M. Baeten, F.W. Vaandrager</td>
<td>An Algebra for Process Creation, p. 29.</td>
</tr>
<tr>
<td>91/31</td>
<td>H. ten Eikelder</td>
<td>Some algorithms to decide the equivalence of recursive types, p. 26.</td>
</tr>
<tr>
<td>91/33</td>
<td>W. v.d. Aalst</td>
<td>The modelling and analysis of queuing systems with QNM-ExSpect, p. 23.</td>
</tr>
<tr>
<td>91/34</td>
<td>J. Coenen</td>
<td>Specifying fault tolerant programs in deontic logic, p. 15.</td>
</tr>
<tr>
<td>Year</td>
<td>Authors</td>
<td>Title</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>92/01</td>
<td>J. Coenen, J. Zwiers, W.-P. de Roever</td>
<td>A note on compositional refinement, p. 27.</td>
</tr>
<tr>
<td>92/02</td>
<td>J. Coenen, J. Hooman</td>
<td>A compositional semantics for fault tolerant real-time systems, p. 18.</td>
</tr>
<tr>
<td>92/03</td>
<td>J.C.M. Baeten, J.A. Bergstra</td>
<td>Real space process algebra, p. 42.</td>
</tr>
<tr>
<td>92/05</td>
<td>J.P.H.W.v.d.Eijnde</td>
<td>Conservative fixpoint functions on a graph, p. 25.</td>
</tr>
<tr>
<td>92/06</td>
<td>J.C.M. Baeten, J.A. Bergstra</td>
<td>Discrete time process algebra, p. 45.</td>
</tr>
<tr>
<td>92/07</td>
<td>R.P. Nederpelt</td>
<td>The fine-structure of lambda calculus, p. 110.</td>
</tr>
<tr>
<td>92/10</td>
<td>P.M.P. Rambags</td>
<td>Composition and decomposition in a CPN model, p. 55.</td>
</tr>
<tr>
<td>92/13</td>
<td>F. Kamareddine</td>
<td>Set theory and nominalisation, Part II, p. 22.</td>
</tr>
<tr>
<td>92/14</td>
<td>J.C.M. Baeten</td>
<td>The total order assumption, p. 10.</td>
</tr>
<tr>
<td>92/15</td>
<td>F. Kamareddine</td>
<td>A system at the cross-roads of functional and logic programming, p. 36.</td>
</tr>
<tr>
<td>92/16</td>
<td>R.R. Seljée</td>
<td>Integrity checking in deductive databases; an exposition, p. 32.</td>
</tr>
<tr>
<td>92/17</td>
<td>W.M.P. van der Aalst</td>
<td>Interval timed coloured Petri nets and their analysis, p. 20.</td>
</tr>
<tr>
<td>92/18</td>
<td>R.Nederpelt, F. Kamareddine</td>
<td>A unified approach to Type Theory through a refined lambda-calculus, p. 30.</td>
</tr>
<tr>
<td>92/19</td>
<td>J.C.M.Baeten, J.A.Bergstra, S.A.Smolka</td>
<td>Axiomatizing Probabilistic Processes; ACP with Generative Probabilities, p. 36.</td>
</tr>
<tr>
<td>92/20</td>
<td>F.Kamareddine</td>
<td>Are Types for Natural Language? P. 32.</td>
</tr>
</tbody>
</table>
92/21 F. Kamareddine

Non well-foundedness and type freeness can unify the interpretation of functional application, p. 16.

92/22 R. Nederpelt
F. Kamareddine

A useful lambda notation, p. 17.

92/23 F. Kamareddine
E. Klein

Nominalization, Predication and Type Containment, p. 40.

92/24 M. Codish
D. Dams
Eyal Yardeni

Bottom-up Abstract Interpretation of Logic Programs, p. 33.

92/25 E. Poll

A Programming Logic for Wo, p. 15.

92/26 T. H. W. Beelen
W. J. J. Stut
P. A. C. Verkoulen

A modelling method using MOVIE and SimCon/ExSpect, p. 15.

92/27 B. Watson
G. Zwaan

A taxonomy of keyword pattern matching algorithms, p. 50.

93/01 R. van Geldrop

Deriving the Aho-Corasick algorithms: a case study into the synergy of programming methods, p. 36.

93/02 T. Verhoeoff

A continuous version of the Prisoner’s Dilemma, p. 17

93/03 T. Verhoeoff

Quicksort for linked lists, p. 8.

93/04 E. H. L. Aarts
J. H. M. Korst
P. J. Zwietering

Deterministic and randomized local search, p. 78.

93/05 J. C. M. Baeten
C. Verhoef

A congruence theorem for structured operational semantics with predicates, p. 18.

93/06 J. P. Veltkamp

On the unavoidability of metastable behaviour, p. 29

93/07 P. D. Moerland

Exercises in Multiprogramming, p. 97

93/08 J. Verhooosel

A Formal Deterministic Scheduling Model for Hard Real-Time Executions in DEDOS, p. 32.

93/09 K. M. van Hee

Systems Engineering: a Formal Approach

93/10 K. M. van Hee

Systems Engineering: a Formal Approach
Part II: Frameworks, p. 44.

93/11 K. M. van Hee

Systems Engineering: a Formal Approach

93/12 K. M. van Hee

Systems Engineering: a Formal Approach
Part IV: Analysis Methods, p. 63.

93/13 K. M. van Hee

Systems Engineering: a Formal Approach
<table>
<thead>
<tr>
<th>Year</th>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>93/16</td>
<td>H. Schepers, J. Hooman</td>
<td>A Trace-Based Compositional Proof Theory for Fault Tolerant Distributed Systems, p. 27</td>
</tr>
<tr>
<td>93/17</td>
<td>D. Alstein, P. van der Stok</td>
<td>Hard Real-Time Reliable Multicast in the DEDOS system, p. 19.</td>
</tr>
<tr>
<td>93/18</td>
<td>C. Verhoef</td>
<td>A congruence theorem for structured operational semantics with predicates and negative premises, p. 22.</td>
</tr>
<tr>
<td>93/19</td>
<td>G-J. Houben</td>
<td>The Design of an Online Help Facility for ExSpect, p. 21.</td>
</tr>
<tr>
<td>93/22</td>
<td>E. Poll</td>
<td>A Typechecker for Bijective Pure Type Systems, p. 28.</td>
</tr>
<tr>
<td>93/23</td>
<td>E. de Kogel</td>
<td>Relational Algebra and Equational Proofs, p. 23.</td>
</tr>
<tr>
<td>93/24</td>
<td>E. Poll and Paula Severi</td>
<td>Pure Type Systems with Definitions, p. 38.</td>
</tr>
<tr>
<td>93/26</td>
<td>W.M.P. van der Aalst</td>
<td>Multi-dimensional Petri nets, p. 25.</td>
</tr>
<tr>
<td>93/27</td>
<td>T. Kloks and D. Kratsch</td>
<td>Finding all minimal separators of a graph, p. 11.</td>
</tr>
<tr>
<td>93/28</td>
<td>F. Kamareddine and R. Nederpelt</td>
<td>A Semantics for a fine λ-calculus with de Bruijn indices, p. 49.</td>
</tr>
<tr>
<td>93/29</td>
<td>R. Post and P. De Bra</td>
<td>GOLD, a Graph Oriented Language for Databases, p. 42.</td>
</tr>
<tr>
<td>93/30</td>
<td>J. Deogun, T. Kloks, D. Kratsch, H. Müller</td>
<td>On Vertex Ranking for Permutation and Other Graphs, p. 11.</td>
</tr>
<tr>
<td>93/31</td>
<td>W. Körver</td>
<td>Derivation of delay insensitive and speed independent CMOS circuits, using directed commands and production rule sets, p. 40.</td>
</tr>
<tr>
<td>Page</td>
<td>Authors</td>
<td>Title</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>93/33</td>
<td>L. Loyens and J. Moonen</td>
<td>ILIAS, a sequential language for parallel matrix computations, p. 20.</td>
</tr>
<tr>
<td>93/34</td>
<td>J.C.M. Baeten and J.A. Bergstra</td>
<td>Real Time Process Algebra with Infinitesimals, p.39.</td>
</tr>
<tr>
<td>93/36</td>
<td>J.C.M. Baeten and J.A. Bergstra</td>
<td>Non Interleaving Process Algebra, p. 17.</td>
</tr>
<tr>
<td>93/38</td>
<td>C. Verhoef</td>
<td>A general conservative extension theorem in process algebra, p. 17.</td>
</tr>
<tr>
<td>93/41</td>
<td>A. Bijlsma</td>
<td>Temporal operators viewed as predicate transformers, p. 11.</td>
</tr>
<tr>
<td>93/42</td>
<td>P.M.P. Rambags</td>
<td>Automatic Verification of Regular Protocols in P/T Nets, p. 23.</td>
</tr>
<tr>
<td>93/43</td>
<td>B.W. Watson</td>
<td>A taxonomy of finite automata construction algorithms, p. 87.</td>
</tr>
<tr>
<td>93/44</td>
<td>B.W. Watson</td>
<td>A taxonomy of finite automata minimization algorithms, p. 23.</td>
</tr>
<tr>
<td>93/48</td>
<td>R. Gerth</td>
<td>Verifying Sequentially Consistent Memory using Interface Refinement, p. 20.</td>
</tr>
</tbody>
</table>
94/01 P. America
M. van der Kammen
R.P. Nederpelt
O.S. van Roosmalen
H.C.M. de Swart

94/02 F. Kamareddine
R.P. Nederpelt

94/03 L.B. Hartman
K.M. van Hee

94/04 J.C.M. Baeten
J.A. Bergstra

94/05 P. Zhou
J. Hooman

94/06 T. Basten
T. Kunz
J. Black
M. Coffin
D. Taylor

94/07 K.R. Apt
R. Bol

94/08 O.S. van Roosmalen

94/09 J.C.M. Baeten
J.A. Bergstra

94/10 T. verhoeff

94/11 J. Peleska
C. Huizing
C. Petersohn

94/12 T. Kloks
D. Kratsch
H. Müller

94/13 R. Seljée

94/14 W. Peremans

94/15 R.J.M. Vaessens
E.H.L. Aarts
J.K. Lenstra

94/16 R.C. Backhouse
H. Doombos

94/17 S. Mauw
M.A. Reniers

The object-oriented paradigm, p. 28.

Canonical typing and Π-conversion, p. 51.


Graph Isomorphism Models for Non-Interleaving Process Algebra, p. 18.


Time and the Order of Abstract Events in Distributed Computations, p. 29.


A Hierarchical Diagrammatic Representation of Class Structure, p. 22.

Process Algebra with Partial Choice, p. 16.

The testing Paradigm Applied to Network Structure, p. 31.


A New Method for Integrity Constraint checking in Deductive Databases, p. 34.

Ups and Downs of Type Theory, p. 9.

Job Shop Scheduling by Local Search, p. 21.

Mathematical Induction Made Calculational, p. 36.

An Algebraic Semantics of Basic Message Sequence Charts, p. 9.
<table>
<thead>
<tr>
<th>Number</th>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>94/18</td>
<td>F. Kamareddine, R. Nederpelt</td>
<td>Refining Reduction in the Lambda Calculus, p. 15.</td>
</tr>
<tr>
<td>94/19</td>
<td>B.W. Watson</td>
<td>The performance of single-keyword and multiple-keyword pattern matching algorithms, p. 46.</td>
</tr>
<tr>
<td>94/20</td>
<td>R. Bloo, F. Kamareddine, R. Nederpelt</td>
<td>Beyond β-Reduction in Church’s λ→, p. 22.</td>
</tr>
</tbody>
</table>