

2IS15 Generic Language Technology (2014-2015)

Construction of NFA and DFA

Assignment 1 (deadline: September 15th 2014, 9:00h)

Introduction

The goal of the first set of assignments of the course Generic Language Technology is to get acquainted with the basic concepts of scanning and parsing. There are several scanner and parser generators, for instance LEX+YACC, that can be used to generate a scanner and parser. In order to get some idea of the underlying technologies, you have to construct a parser "by hand".

The first exercise focusses on the generation of scanners. Developing a scanner by hand is quite tedious and error prone. In this exercise you have to use two libraries, one library provided by Java "RegEx" to generate NFAs and one library "Automaton" to generate DFAs.

Documentation on the Java "RegEx" library can be found at <http://docs.oracle.com/javase/tutorial/essential/regex/> and <http://docs.oracle.com/javase/8/docs/api/java/util/regex/Pattern.html>. Documentation on the "Automaton" library can be found at: <http://www.brics.dk/automaton/>. It may also be useful to consult the "Frequently Asked Questions" page: <http://www.brics.dk/automaton/faq.html>. The syntax of the regular expressions can be found at: <http://www.brics.dk/automaton/doc/index.html?dk/brics/automaton/RegExp.html>.

Setup

1. Download the file `assignment1.zip` from <http://www.win.tue.nl/~mvdbrand/courses/GLT/1314/exercises/assignment1.zip>.
2. This zip file contains an Eclipse example project how to use these libraries.

Generating Scanners based on (Simple) Lexical Definitions

1. Implement the following lexical definitions.

```
ID ::= [a-z] [a-z0-9]*  
NAT ::= [0] | [1-9] [0-9]*
```

Give a list of input strings for testing, make sure it also contains a number of input strings that will not be accepted.

2. Create a lexical definition for recognizing string representations of the form "abc". Make sure that double quotes are properly escaped. Take into account double quotes and other escaped characters appearing within such string representations. Give a list of input strings for testing, make sure it also contains a number of input strings that will not be accepted.
3. At [http://en.wikipedia.org/wiki/Comment_\(computer_programming\)](http://en.wikipedia.org/wiki/Comment_(computer_programming)) the comment conventions of a number of programming languages are given. Implement the C and Java comment plus one extra. Give a list of input strings for testing, make sure it also contains a number of input strings that will not be accepted.

Submission

Submit the following file via PEACH:

1. `RegexText.java`, this Java module should contain code for each of the required regular expressions for NFA and DFA plus the corresponding tests. The tests should generate sufficient output such that the tests can be reviewed based upon the generated output.