
Rascal Requirements and Design Document

Paul Klint
Tijs van der Storm
Jurgen Vinju

Table of Contents

Rationale	2
Introduction	2
Rascal for various audiences	3
Generic arguments	3
Rascal for ASF+SDF programmers	4
Rascal for imperative and object-oriented programmers	5
Requirements	6
Rascal at a glance	7
Modules	7
Names	8
Types and Subtypes	8
Annotations	10
Variables	10
Functions	11
Patterns	13
Expressions	16
Statements	17
Failure and side-effects	18
Examples	18
Hello world	18
Table of squares	19
Booleans	19
Abstract Graph datatype	21
Tree traversal	22
Substitution in Lambda	25
Renaming in Let	26
Renaming in Let using globals	27
Concise Pico Typechecker	28
Pico evaluator	29
Pico evaluator with globals	31
Pico control flow extraction	33
Pico use def extraction	34
Pico uninitialized variables	34
Pico common subexpression elimination	34
Pico constant propagation	35
Pico Reaching definitions	36
Structured lexicals: numbers	37
Structured Lexicals: strings	37
Symbol table with scopes	38
Innerproduct	39
Bubble sort	40

Generic Bubble sort	40
Read-Eval-Print Loop (REPR) [Needs further discussion]	41
The Rascal standard library	42
Main functions	42
Additional functions on lists sets, maps and relations	43
Other functions	44
Source code	44
Syntax Definition	58
Prototyping/implementation of Rascal	58
Rascal implementation ideas	59
Data structures	60
Interpreter	60
Compiler	61
Issues	62
Graveyard	62
Mapping features to datatypes	62
Visitor definitions (UNDECIDED and INCOMPLETE)	63
Outdated examples	65
Generating Graph files in Dot format	65
Integration with Tscripts (Outdated)	65
Introduction	65
Requirements	65
Different styles of Type Declarations	65
Global Flow of Control	66
Modularization	67

Note

This document is a braindump of ideas that is slowly converging to a coherent design. See the section called “*Issues*” [62] for the issues that have to be resolved.

Rationale

In the domain of software analysis and transformation, there exists a phlethora of domain-specific languages for defining grammars, rewrite rules, software analysis and the like. So why embark on the design of yet another DSL in this area? We see the following arguments for this:

- Many existing DSLs are based on more or less exotic concepts that do excite researchers but are frightening for users without an appropriate research background.
- The notation used in many DSLs is uninviting to say the least.
- The scope of DSLs is usually narrow (this is where the word "domain-specific" kicks in). The tasks involved in carrying out a code analysis or renovation project require several DSLs. Integration between these DSLs is insufficient.
- As designers of various DSLs (ASF+SDF, Tscript, Rscript, ...) we have seen the positive as well as the negative side of DSLs. We certainly know howto implement DSLs in this area.
- We see an opportunity for a user-friendly, conceptually high-level, and rich DSL for all tasks related to software analysis and transformation.

With this background and an essential dose of optimism, we embark on this trip

Introduction

The goals of the envisaged language (with working name Rascal) are:

- Providing a successor of ASF+SDF that has of all its benefits and fixes all of its shortcomings.
- Separating pure syntax definitions (SDF) from function definitions.
- Easy syntax-directed analysis of programming languages.
- Easy fact extraction.
- Easy connection of fact extraction with fact manipulation and reasoning.
- Easy feedback of analysis results in source code transformation.
- Efficient and scalable implementation.
- Unsurprising concepts, syntax and semantics for a wide audience. Where possible we will stay close to C and Java notation.

Many of the above goals are to a certain extent already met in the current design of ASF+SDF, and the current design of RScript. What is missing is the connection (and to be honest: an efficient implementation of relational operators). Alas, any bridge between the two languages is both complex to manage and an efficiency bottleneck. This work is an attempt to consolidate this engineering trade-off. This basically means that we include most features of the RScript language into ASF+SDF. Although we take these languages as conceptual starting point, Rascal is a completely new design that has an imperative semantics at it's core rather than a functional semantics. As a whole, Rascal is a simpler but more expressive language.

Rascal for various audiences

In this section we enumerate numerous facts about Rascal that advertise it to different audiences

Generic arguments

What is good about Rascal in a few words?

- Rascal is a DSL for source code analysis and transformation. It provides a plethora of high level statements and expressions, taking away the boilerplate of implementing and debugging tools that manipulate programs.
- Rascal combines the best features of imperative programming with the best features of functional programming and term rewriting.
 - Simple structured statements for control flow and variable assignments for data flow are powerful and simple features of the imperative programming paradigm. They allow control flow and data flow to be understandable and traceable.
 - From functional programming we borrow that all values are immutable and non-null. Issues with aliasing and referential integrity, such as frequently occur in imperative and OO programming therefore do not exist in Rascal.
 - The Rascal type system is as powerful as most functional languages (higher-order polymorphic functions), however to make the language debuggable and understandable it, in principle, does not provide type inference.
 - From term rewriting we inherit powerful pattern matching facilities, integration with context-free parsing and concrete syntax.
- Rascal supports both a scripting experience, and a compiled program experience.
- Rascal is type safe, but flexible. It's type system prevents common programming errors, but still allows ample opportunity for reusable code. The reasons are the we allow co-variance in the sub-typing relationship, high-order polymorphic functions and parameterized data-types.

- Rascal allows different styles of programming. From extremely high level specification, down to straight imperative programming.
- Rascal was inspired by and borrows from several other DSL's for program analysis and transformation in academia and industry, namely ASF+SDF, Rscript, TXL, TOM, DMS, Stratego, Elan, Maude, Grok, Haskell, ML and Setl.
- Rascal integrates seamlessly with Eclipse IMP and The Meta-Environment.

Rascal for ASF+SDF programmers

Rascal is the successor of ASF+SDF. What's the difference? What's the same?

- Rascal has roughly all the high level features of ASF+SDF and some more. Old ASF+SDF specifications can be transformed to Rascal programs using a conversion tool.
- Rascal still uses SDF for syntax definition and parser generation.
- Rascal has a module system that is independent of SDF. Rascal modules introduce a namespace scope for variables and functions, which can be either private or public. Rewrite rules are global as in ASF+SDF. Modules can have type parameters as in SDF, which are instantiated by import statements.
- In Rascal, patterns and variables in concrete syntax may optionally be quoted and escaped, and support explicit declaration of the top non-terminal to solve ambiguity.
- Rascal rules read in the order of execution instead of first the left-hand side, then the conditions, and then the right-hand side of ASF+SDF equations
- Rascal has primitive and efficient implementations for sets, relations and maps
- Rascal can be used without SDF, supporting for example regular expressions and abstract data types (pure ATerms)
- Rascal has primitive support for functions, which have a fixed syntax, always return a value and have imperative control flow statements. Adding a function will not trigger the need for regenerating parse tables. Function types can be polymorphic in their parameters and also allow functions as arguments to implement reusable algorithms.
- The imperative nature of Rascal allows you to factor out common code and nest conditionals, unlike in ASF+SDF where alternative control flow paths had to be encoded by enumerating equations with non-overlapping conditions.
- Rascal is an imperative language, which natively supports I/O and other side-effects without work-arounds. When backtracking occurs, for example over list matching, Rascal makes sure that most side-effects are undone, and that I/O is delayed until no more backtracking can occur. Even rewrite rules support side-effects in Rascal.
- Rascal has native support for traversals, instead of the add-on it used to be in ASF+SDF. The visit statement is comparable to a traversal function, and is as type-safe as the previous, and more programmeable.
- Instead of accumulator values of traversal functions in ASF+SDF, Rascal simply supports lexically scoped variables that can be updated using assignments.
- Rascal adds specific expressions for relational calculus, all borrowed directly from RScript.
- When programming using Rascal functions, instead of rules, the control flow of a program becomes easily traceable and debuggable. It is simply like stepping through well structured code.
- Rascal is based on a Java interpreter, or a Java run-time when compiled. So the code is more portable.

Rascal for imperative and object-oriented programmers

Rascal is an imperative DSL with high level statements and expressions specifically targeted at the domain of analysis and transformation of source code:

- Rascal is safe: there are no null values, and all values are immutable. Source code and abstract syntax trees, and the facts extracted from them are immutable. The Rascal interpreter and compiler make sure this is implemented efficiently. Without mutability it is easy to combine stages of your programs that analyse or annotate with stages that transform. Sharing a value does not introduce a coupling like in OO, simply because changes are only visible to the code that changes the values.
- Rascal is extra safe: it has a type system that prevents casting exceptions and other run-time failures. Still the type system specifically allows many kinds of combinations. For example, unlike in Java a set of integers is a subtype of a set of numbers (co-variance), which allows you to reuse algorithm for sets of numbers on sets of integers. It also provides true polymorphic and functions (no erasure), and functions can safely be parameters to other functions.
- Rascal provides high level statements and expressions for:
 - Visitors in all kinds of orders, expressed very concisely, and type safe.
 - Pattern matching and construction (with concrete syntax!)
 - Equation/constraint solving
 - Relational calculus
 - Rewrite rules for normalization/canonicalization of any kind of data-structure
 - Support for parsing using context-free grammars (via importing modules from the SDF language).
 - (de)Serialization of values
 - Communication with databases
- Rascal provides typed data constructors for common mathematical structures, such as:
 - terms (a.k.a. abstract data types, tree nodes)
 - parse trees (derivations of context-free grammars, for concrete syntax and direct manipulation of source code)
 - relations
 - sets
 - maps
 - graphs
 - tuples
- In Rascal you can implement high-fidelity source-to-source transformations. Without too much overhead, programs can do extensive rewriting of the source code without the loss of particular layout standards or source code comments.
- Rascal is syntax-safe. When you use Rascal to generate or transform source code, it statically detects whether the resulting source code is syntactically correct.

- Rascal is executed by an interpreter written in Java, or it can be compiled to Java classes.

Requirements

- R1: Runtime speed: large-scale analysis of facts is expensive (frequently high-polynomial and exponential algorithms). A factor speedup can mean the difference between a feasible and an unfeasible case.
- R2: Old ASF+SDF programs are translatable to Rascal.
- R3: Edit Rascal and SDF and compile complex programs within a few minutes maximally (parsetable generation is a major bottleneck in current ASF+SDF. This needs to be fixed.)
- R4: Concrete syntax: for readability and easy parsing of a wide range of source languages.
- R5: **(dropped)** Functional (no side-effects), dropped due to incompatibility with R7, R6
- R6: File I/O (contradicts R5).
- R7: Easily accessible fact storage (similar to a heap, but remember R5 and the details of backtracking).
- R8: List matching (because of R2, influences R7, but also very handy for manipulating lists in concrete syntax).
- R9: Nesting of data-structures: relations can be nested to model nested features of programming languages (such as scoping), allowing to factor out common code.
- R10: Syntax trees can be elements of the builtin data-structures (but not vice versa).
- R11: Try to keep features orthogonal: try to keep the number of ways to write down a program minimal, this is not a law since other requirements take precedence
- R12 **(new)**: Minimize possible syntactic ambiguities; resolve them by type checking.
- R13:**(new)**: Integrates well with refactoring infra-structure (i.e. can provide appropriate interfaces with pre-condition checking, previews and commits as found in interactive refactoring contexts)
- R14 **(new)**: no 'null' values, preventing common programming errors
- R15 **(new)**: all values immutable, preventing common programming errors and allowing for certain kinds of optimizations
- R16 **(new)**: should be able to match and construct strings using regular expressions (for making the simpler things simple, if you can do without a grammar, why not?)
- R17 **(new)**: can get/set data from databases, such as the pdb from Eclipse IMP, but possibly also from ODBC/JDBC data sources.
- R18 **(new)**: type safe, but flexible. We want a type system that prevents common programming errors, but still allows ample opportunity for reuse.
- R19 **(new)**: syntax safe, programmers should not be allowed to construct programs that are syntactically incorrect w.r.t a certain context-free grammar.
- R20 **(new)**: backtracking safe, programmers should not have to deal with the mind boggling feature interactions between side-effects and backtracking.
- R21 **(new)**: traceable/debuggable, programmers should be able to easily trace through the execution of a Rascal program using the simplest of debugging tools, like printf statements, and the use of a simple debugging interface which allows to step through the source code and inspect values in a transparent fashion.

- R22 (**new**): minimize the use of type inference, such that the programmer must always declare her intentions by providing types for functions, data-types and variables. This makes debugging easier and providing clear error messages too. When variables are implicitly bound by pattern matching or related functionality, exceptions to this requirement might be made in favor of conciseness.
- R23 (**new**): allow the implementation of reusable modules and functions (i.e. parametric polymorphism and or functions as parameters).
- R24 (**new**): we need something like rewrite rules for implementing data-types that are always canonicalized/normalized. For some analysis algorithms this allows the programmer to implement domain specific optimizations over plain relational calculus or tree visiting that actually needed for scalability.

Rascal at a glance

Rascal consists of the following elements:

- Modules to group definitions, proving scopes and visibility constructs
- A type system and corresponding values, providing parameterized types, polymorphic functions and higher-order parameters.
- Variables to associate a name with a value in some scope.
- Parameterized functions.
- Abstract, regular expression and syntax patterns to deconstruct (match) values and to construct (make) them.
- Expressions provide the elementary computations on values.
- Statements for providing structured control flow and more advanced control flow in computations, such as visitors and fixed point computations

These elements are summarized in the following subsections.

Modules

Modules are the organizational unit of Rascal. They may:

- Import another Rascal module using `import`. We do not allow circular imports.
- Extend another Rascal module using `extend`. This includes a verbatim copy (similar to `#include`) of the-module-to-be-extended in the current module. We allow redefinitions of declared names. We do not allow circular extension.
- Import SDF modules using `import-syntax`.
- Import Java modules (for the benefit of functions written in Java) using `import-java`.
- Define datatypes, subtypes, or rules and functions.
- Be parameterized with the names of formal types that are instantiated with an actual type when the module is imported.
- Contain a main function that is the starting point of execution:

```
public int main(list[str] argv) { ... }
```

Functions, subtypes and datatypes may be either private to a module or public to all modules that import the current module. Rules are always public and are globally applied.

Modules introduce a namespace and qualified names (using the `::` operator) may be used to uniquely identify elements of a module from the outside. Inside the module, this qualification is implicit.

Caution

We should clarify what we want to have:

- modules: yes
- a nested structure (directory style) of modules: yes
- packages: no?

Names

Rascal aims at seamless integration with Java and its names adhere to the following conventions:

- A Rascal Name is identical to a Java Identifier. All reserved words in both Java and Rascal cannot be used as a Name. *For better syntax errors it is probably better to warn for Java keywords later on.*
- Sorts and Symbols are inherited from SDF but we extend sort (that may only start with an uppercase letter in SDF) to be a more liberal Name. *SDF Sorts may include a dash symbol (-) which is not allowed in Java Identifiers. Forbidding the dash requires changing all existing SDF definitions. Allowing a dash in Names requires an added check in the compiler that Identifiers with dashes are generated as Java names. Suggestion: add dash and check in compiler.*

Types and Subtypes

Types

The type system (and notation) are mostly similar to that of Rscript, but

- We have a type hierarchy that defines a partial order on types.
- There is a single root for this type hierarchy, it is called "value".
- There are built-in types (`bool`, `int`, `double`, `str`, `loc`, `void`).
- Symbols (as defined by an SDF module) are also types.
- All syntactic types are a subtype of the type `tree` that corresponds to UPTR. Up casts from a subtype to an enclosing type are automatic; Down casts require a run-time check/cast. The type `tree` is "special" in the following sense:
 - Parsers generate values of type `tree`.
 - Although the type `tree` can be defined in Rascal, its definition is built-in in order to preserve the consistency with the parser.
- Types for sets, maps, relations and tuples can be formed from types; Maps, relations and tuples can have optional column names.
- Functions can be polymorphic in their parameters as well as in their return type.
- Function parameters can be function types, meaning that the name of a properly typed function can be passed as a parameter.
- The last argument of a function may be of the form `list[T] Name . . .` and indicates a variable list of remaining arguments of type `T`.

- Arbitrarily nested expressions may contain calls to overloaded functions that need to be resolved.
- Data declarations may introduce new structured types.

Note

We assume that all constructors for a datatype lead to a corresponding function to construct a term of that datatype. This eliminates the need to quote abstract terms in statements.

Note

Constructor names of datatypes may be quoted in order to avoid clashes with reserved words or illegal names.

- Types may include type variables like $\&T$ as in Rscript with the following refinement:
 - In the type `rel[$\&T$]`, the type variable $\&T$ is bound to the tuple type of the relation, i.e. an actual type `rel[int, str]` will bind $\&T$ to the type `tuple[int, str]`.
 - In a similar fashion, in `map[$\&T$]`, the type variable $\&T$ is bound to the tuple type of the map.
- A type declaration introduces a name for a new type that is a subtype of a given type, e.g.,

```
type rel[node,node] Graph
```

introduces the type `Graph` as a subtype of `rel[node, node]`. There are no built-in subtypes.

- The name that is introduced by a type declaration may even be parameterized with one or more type variables, e.g.,

```
type rel[&Node, &Node] Graph[&Node]
```

Of course, the type variables that are used in the type in the left part should occur as parameters in the right part of the definition.

Subtypes

The type hierarchy leads to a subtype relation $<:$ that can be defined (in ASF+SDF) as follows:

```
%% Subtyping rules (<:) on Types

[let01] T <: value                                = true
[let02] T <: T                                      = true
[let03] T <: & N                                    = true
[let04] T <: &N <: T'                              = T <: T'
[let05] N[&T] <: N[&U]                             = true
[let06] list[T] <: list[T']                         = T <: T'
[let07] set[T] <: set[T']                           = T <: T'
[let08] tuple[ONTS] <: tuple[ONTS']                 = ONTS <: ONTS'
[let09] map[T1, T2] <: map[T1', T2']                = T1 <: T1' & T2 <: T2'
[let10] map[T1, T2] <: rel[T1', T2']                = T1 <: T1' & T2 <: T2'
```

```

[let11] rel[ONTs] <: rel[ONTs']           = ONTs <: ONTs'

[let12] rel[ONTs] <: set[tuple[ONTs']]     = ONTs <: ONTs'

[let13] set[tuple[ONTs]] <: rel[ONTs']     = ONTs <: ONTs'

[let14] T <: T' == true,                  %% covariant
      ONTs' <: ONTs == true              %% contravariant
      =====
      fun T N (ONTs) <: fun T' N (ONTs') = true
[default-let]
      T <: T'                            = false

```

Here T , T_1 , T_2 , T' etc represent types, and $ONTs$, $ONTs'$ etc represented lists of OptionallyNamedTypes, e.g., a type that optionally followed by a named. We omit the definition of \leq on OptionallyNamedTypes. The above rules also describe certain equivalences between types, e.g., a map can be compared with a relation, a relation can be compared with a set of tuples, and the like.

The subtyping relation for function types is tricky and treats result type and argument types differently. We have, for instance, that

```

fun set[int] f(int i) <: fun set[&T] f(int i)  %% since int < &T
fun int f(number x) <: fun number f (int x)    %% assuming int <: number

```

but *not* that

```

fun number f(int x) <: fun number f (number x) %% assuming int <: number

```

In addition to the above rules for subtyping, each type definition introduces an additional subtype relation, e.g., the previous definition for `Graph`,

```

type rel[node,node] Graph

```

introduces the following subtype relation:

```

Graph <: rel[node,node]

```

Caution

How do we handle the optional names that occur in types? Currently, we throw them away in the subtype relation. Should we also compare these names?

Annotations

Annotations may be associated with any value and are represented by a mapping of type `map[str,value]`, i.e., annotation names are strings and annotation values are arbitrary values of type `value`. The standard library provides functions for handling annotations.

Variables

Variables are names that have an associated scope and in that scope they have a value. A variable declaration consists of a type followed by the variable name and---depending on the syntactic position---they are followed by an initialization. There are no null values, which implies that all variables must be initialized at declaration time. Also, this implies that all expressions must return a value. Especially for functions, this means that all execution paths of a function must have a return statement.

Variables may be introduced at the following syntactic positions:

- As formal parameters of a function. Their scope is the function and they get their initial value when the function is called.
- Local variables in a function body are declared and initialized. Their scope is the function body.
- Variables in patterns. For patterns in match positions, variables are initialized during the match and their scope is the rule in which they occur. For patterns in make positions, the values of variables are taken from the local scope.
- For variables introduced by pattern matching in conditional statements (if-then, and while), if the condition succeeds, the scope of the variables are the block of code that is executed conditionally.
- Variables in anti-patterns are never visible, but nevertheless their names are reserved in the scope that they would have had when the pattern was a normal positive matching pattern.
- Variables that are introduced by generators in comprehensions or for statement, have the comprehension, respectively, for statement as scope.
- Global variables are declared and ALWAYS initialized at the top level of each module and **these initialization may not contain function calls**. Functions that use a global variable have to be explicitly declare it as well. The value of a global variable can be used and replaced by all functions that have locally declared it.

We will see below that there are certain contexts in which assignments to variables are undone in the case of failure.

Functions

Overview

A function declaration consists of a visibility declaration, the keyword `fun`, result type, a function name, typed arguments and a function body. Functions without a result type have type `void`.

A visibility declaration is one of the keywords `public` or `private` (default).

A function body is a list of statements, optionally separated by semi-colons. Each unique control flow path through a function must have a return statement, such that each function always returns a proper value.

Functions can raise exceptions and these may -- for documentation purposes -- optionally be declared as part of the function signature (i.e., we have unchecked exceptions in contrast with checked exceptions in Java).

Functions preceeded by the keyword `"java"` have a body written in Java. They have the following properties:

- Arguments and result are pure Rascal values.
- Java functions cannot acces the global state of the Rascal program. Their only interface is via input parameters and a result value. They cannot access Rascal global variables.
- Side effects caused by Java functions in the Java state, are not undone in the case of backtracking.

Note

These Java functions are primarily intended for implementing library functions and for interfacing with Eclipse.

Functions preceeded by the keyword `"memo"` are memo functions that cache previous arguments/result combinations.

Caution

Do we need library functions to tune the memo behaviour (e.g., flush cache, change cache size)?

Overloading

Functions may be *overloaded*, i.e., functions with the same name but with different argument and result types can be declared. Overloading is subject to the restriction that if several versions of a function are declared with the same name, then one or more of the following conditions should hold:

- the number of arguments differs;
- there is at least one argument position with incomparable types w.r.t. the type hierarchy, i.e., not ($A <: B$) and not ($B <: A$).

Here are some examples.

Here are, after all the warnings, the original examples:

```
%% OK, since int and double are incomparable

fun int +(int a, int b) { ... }
fun double +(double a, double b) { ... }

%% KO, since int is comparable to value
%% (int is -- like all types -- a subtype of value)

fun int +(value a, value b) { ... }
fun int +(int a, int b) { ... }

%% OK since for every two definitions selected from this
%% set there is at least one parameter position different;
%% in this example the overloading on result types will never
%% be used to select the function

fun int    +(int a, int b)
fun double +(int a, double b)
fun double +(double a, int b)
fun double +(double a, double b)

%% KO, since arguments cannot be resolved.

fun int randomValue()
fun str randomValue()
```

How should overloading be resolved in the presence of type parameters? Our approach is simple but effective: a type parameter $\&T$ is taken to be of type `value`, unless it is explicitly constrained to be smaller than (or smaller than or equal to) another type U by writing $\&T <: U$. For the sake of overloading resolution, a constrained type parameter will be considered to have the type of the constraint, i.e., in `fun F(&T < U)` we assume the $\&T$ to be of type U .

Here are more examples:

```
%% KO, because &T might be any type (i.e. value),
%% which is comparable to int

fun str f(int x)
fun str f(&T x)
```

```
%% KO, since both T's may bind to comparable types and
%% value and int are comparable

fun int f(&T x, value y)
fun int f(&T y, int x)

%% OK, since the second arguments have incomparable types

fun int f(&T x, int y)
fun int f(&T y, double x)
```

Formal description of overloading

Given the subtype relation \leq defined earlier, we can easily formalize overloading resolution by defining two predicates: `incomparable` describes when two lists of `OptionallyNamedTypes` can be compared and `may-overload` describes when two function types satisfy the overloading restrictions:

```
%% Incomparable

[com1] incomparable(ONTs, ONTs') =
    not((ONTs <: ONTs') | (ONTs' <: ONTs))

%% May-overload

[mo-1] may-overload(fun T N (ONTs), fun T' N' (ONTs')) =
    not(N == N') |
    incomparable(ONTs, ONTs')
```

Anonymous Functions

We also allow *anonymous functions*, i.e., functions that are declared locally and can be passed as argument to another function, be returned as value of a function, or even be stored as value in a set or relation. When such functions are called they are called in the lexical scope in which they were defined. For instance, the Rascal standard Library defines a `mapper` function that applies a function to a list or set. An anonymous function can be used to define such function arguments. Here is a function `addOne`, that adds 1 to each element of its argument list:

```
fun list[int] addOne(list[int] L)
    return mapper(L, fun int (int N) { return N + 1 })
```

Patterns

We distinguish four kinds of patterns:

- *Abstract* patterns: prefix dataterms that are generated by a signature.
- *Regular expression* patterns: conventional regular expressions
- *Syntax* patterns: textual fragments that are generated by a context-free grammar.
- *Lexical* patterns: a special case of syntax patterns dealing with lexical notions like identifiers, numeric constants, and the like.

Patterns may contain variables and can occur in two syntactic positions:

- *Match* positions where the patterns is matched against another term and the variables in the pattern are bound when the match is successful. Examples of match positions are:

- In a case construct, immediately after the `case`, `acase` or `rcase` keyword.
- The left-hand side of a `rules` or `arules` statement.
- In a generator, where generated values are matched against the pattern.
- *Make* positions where the pattern is used to construct a new term (after replacing any variables in the pattern by their values. Examples of make positions are:
 - In a case construct, immediately after the `=>` operator.
 - The right-hand side of a `rules` or `arules` statement.
 - As an ordinary expression in the form of a quoted syntax pattern, or a call to a constructor for a `dataterm`.

A pattern may be turned into an *anti-pattern* by prefixing it with the symbol `!`. An anti-pattern matches in all cases where the original pattern does not match. A match of an anti-pattern cannot bind any variables but these variables are nonetheless reserved in the corresponding scope.

Abstract Patterns

Datatype declarations introduce a signature of abstract terms. These terms (possibly including typed variables as introduced for concrete patterns) may be used as abstract patterns at the same position where concrete patterns are allowed. Type declarations define an inclusion relation between types.

Regular Expression Patterns

Regular expression patterns are ordinary regular expressions that are used to match a string value and to decompose it in parts and also to compose new strings. In a match position, a regular expression may be preceded by a newly declared variable (always of type `list[str]`) and the symbol `:=`. When the variable is present it will contain the matches obtained by the regular expression.

Syntax Patterns

There is a notation of a *syntax pattern*: a (possibly quoted) concrete syntax fragment that may contain variables. We want to cover the whole spectrum from maximally quoted patterns that can unambiguously describe **any** syntax fragment to minimally quoted patterns as we are used to in ASF+SDF. Therefore we support the following mechanisms:

- Optionally typed variables, written as `<TYPE NAME>` or `<NAME>`.
- Quoted patterns enclosed between `[|` and `|]`. Inside a fully quoted string, the characters `<`, `>` and `|` can be escaped as `\<`, `\>`, `\|`. Fully quoted patterns may contain variables.
- Unquoted patterns are an (unquoted) syntax fragment that may contain variables.

Quoted and unquoted patterns form the syntax patterns that are supported in Rascal.

Examples are:

- Quoted syntax pattern with typed variables:

```
[ | while <EXP Exp> do <{STATEMENT ";" }* Stats> od | ]
```

- Quoted syntax pattern with untyped variables:

```
[ | while <Exp> do <Stats> od | ]
```

- Unquoted syntax pattern with typed variables:

```
while <EXP Exp> do <{STATEMENT ";"}* Stats> od
```

- Unquoted syntax pattern with untyped variables:

```
while <Exp> do <Stats> od
```

Obviously, with less quoting and type information, the probability of ambiguities increases. Our assumption is that a type checker can resolve them.

Note

Implementation hint (used to check the examples in this document). For every sort S in the syntax definition add the following rules:

```
 $S$                                 -> Pattern
"<" "S"? Name ">"                ->  $S$ 
```

Lexical Patterns

A special case of syntax patterns, are lexical patterns that describe lexical notions such as identifiers, numeric constants and the like. Lexical patterns can appear as part of a concrete syntax pattern. Our solution is a middle road between the original solution used in ASF+SDF (simple but not type safe) and the current solution in ASF+SDF (complex but type safe). The key idea is that a constructor function is implicitly created for every lexical definition, i.e., when the SDF definition defines the lexical sort LS , then the following lexical constructor function ls (the sort name in lower case) is implicitly defined:

```
ls( LEXARGS ) -> LS
```

LEXARGS consist of zero or more:

- String constants.
- Typed Variables enclosed in angle brackets (as used in other patterns).

The LEXARGS argument combined should form a strings of sort LS .

Note

The current solution in ASF+SDF requires that all intermediate lexical constructors are explicitly written in the pattern. Here we relax this requirement.

For instance, given a lexical syntax for numbers (part of an SDF definition):

```
sorts DIGIT NAT-CON
lexical syntax
  [0-9]    -> DIGIT
  DIGIT+   -> NAT-CON
```

A rule that would remove leading zeros looks like this:

```
natcon("0" <DIGIT+ Ds>) => natcon(<Ds>)
```

Note

Unlike the ASF+SDF solution, it is not necessary to make a distinction between ordinary variables, lexical variables, and layout variables. Here, all variables are treated equal and their syntactic position determines how they are used.

We also allow arbitrary character classes as type in a pattern variable, e.g., one could also write:

```
natcon("0" <[0-9]+ Ds>) => natcon(<Ds>)
```

Expressions

Expressions correspond roughly to Rscript expressions with some extensions:

- There are lists, sets and relations together with comprehensions for these types.
- Generators in comprehensions may range over syntax trees.
- Generators in comprehension may range over integer ranges of the form $[B..E]$ or $[B, I, ..E]$, where B indicates the begin value, E the end value, and I the optional increment between values.
- Generators may have a strategy option to indicate:
 - `top-down`
 - `top-down-break`
 - `bottom-up` (this is the default)
 - `bottom-up-break`

The two other strategy option (`innermost` and `outermost`) are only meaningful in the context of a `visit` statement.

- The complete repertoire of operators in Rscript is available but we have applied some rationalization. Note that with each operator, we also associate a named function that can be used as function argument.
 - `+` is now used for addition, union and concatenation (the operator `union` is gone).
 - `-` is now used for subtraction, difference (the operator `\` is gone)
 - `*` is now used for multiplication.
 - `/` is used for integer and double division.
 - `&` is used for intersection
- There are new assignment operators `+=`, `-=`, `*=`.
- The relational operators `==` (`equal`) and `<` (`less`) are available for values of all types. They are used to define `<=` (`less_equal`), `>` (`greater`) and `>=` (`greater_equal`) as well as the functions `min` and `max`.
- For a binary relation R that is a map (i.e. it associates a single value with each domain element) $R[N]$, returns the single image value corresponding with N . This expression may also occur as left-hand side of an assignment:

```
 $R[N] = V$ 
```

first removes from R the tuple with domain value N and then adds the tuple $\langle N, V \rangle$. The net effect is that a new map value is assigned to R .

- An index operator $R[N]$ can be applied to values of types list, set, tuple, rel, string and returns the N -th element.
- The Rscript image notation $R[-, n]$ is discontinued.

Caution

Do we need a cast operator to down cast values?

Statements

Rascal has the following statement types, which can be nested and composed in the usual structured manner:

- Variable declaration with initialization.
- An assignment statement assigns a value to a variable.
- If-then statement and if-then-else statement.
- A `return` statement returns a value from a function, or just returns (for functions with result type `void`). Note that return jumps out of an entire function, even if it is nested in a complicated control flow statement such as `visit`.
- A `yield` statement that delivers a replacement value during a traversal initiated by a `visit` statement, or a `switch` statement.
- A `fail` statement which jumps to the deepest nested choice point (i.e. a `switch` or a `visit`) and tries the next option available. Fail is the dual of `return/yield`.
- A `throw` statement can raise an exception.
- A `switch` statement is similar to a switch statement in C or Java and for a given subject term, it corresponds to the matching provided by the left-hand sides of a set of rewrite rules. However, it provides **only** matching at the top level of the subject term and does not traverse it. The type of each pattern must be identical to the type of the subject term (or be a subtype of it). It is an error if no case matches.
- A rewrite rule (not a top-level statement, but a child of `switch` and `visit`) consists of a `Pattern` followed by `:` and a statement that returns the replacement value:

```
case Pattern : Statement
```

A statement may consist of declarations and statements and is implicitly extended with a `fail` statement. The statement must therefore yield or return a value or the case as a whole will fail. To maintain some resemblance with rewrite rules, we also support the form

```
case Pattern => Replacement
```

which is an abbreviation for

```
case Pattern : yield Replacement
```

Depending on the kind of pattern we use a different keyword:

- `case` for (concrete) syntax patterns.
- `acase` for abstract patterns.
- `rcase` for regular patterns.

Note

These three different keywords are not very elegant but they help to remove disambiguities.

- A `visit` statement corresponds to a traversal function from ASF+SDF. Given a subject term and a list of rewrite rules it traverses the term. Depending on the precise rules it may perform replacement (mimicking a transformer), update local variables (mimicking an accumulator) or a combination of these two. The `visit` statement may contain the same strategy options as a generator and also:

- `innermost` = compute a fixed-point: repeat a bottom-up traversal as long as the traversal function changes values.
- `outermost` = compute a fixed-point: repeat a traversal traversal as long as the traversal function changes values.
- A `for` statement to repeat a block of code.
- A `solve` statement to solve a set of linear equations.
- A `try` statement can be used to execute a statement and to catch any exception raised by that statement.

Failure and side-effects

There are two contexts in which side-effects, i.e., assignment to variables, have to be undone in case of failure. These contexts are a rewrite rule in a switch or visit statement. If the pattern on the left-hand side of the rule matches there are various possibilities:

- All control flow path through the right-hand side of the rule end in a return statement. In this case, the rule can not fail and all side-effects caused by the execution of the right-hand side are committed.
- One or more control path can fail. This can be caused by an explicit fail statement or an if-then statement with missing else-branch. In the case of failure all side-effects are undone.

Caution

Will we undo all side-effects or only the side-effects on global variables?

- If a rule fails there are two possibilities:
 - the left-hand side contains a list pattern that has more matching options; the next option is tried.
 - the left-hand side contains a list patterns that has no more matching options or it contains no list pattern at all; the next rule is tried.

Examples

Here we list experimental examples of Rascal code to try out features.

Hello world

The ritual first example:

```
module Hello

public fun int main(list[str] argv) {
    println("Hello, this is my first Rascal program");
    return 0;
}
```

Caution

How do we type `main`? This is the type as defined it earlier but this is verbose. Can we provide alternatives including `void main()`, this would eliminate the need for the ugly return statement:

```
module Hello
```

```
public fun void main()
  println("Hello, this is my first Rascal program")
```

Table of squares

Another ritual example, printing a table of squares:

```
module Squares

public fun int main(list[str] argv){
  int N = toInt(argv[0]);
  map[int,int] squares = {};
  for(int I : [1..N]){
    squares[I] = I * I
  }
  println("Table of squares from 1 to ", N);
  println(squares);
}
```

Caution

We need a mechanism to insert values in strings, e.g.

```
println("Table of squares from 1 to <N>");
```

Booleans

It seems that every language specification effort has to produce a specification of the Booleans at some moment, so let's try it now. We try several variations.

We use the following common syntax:

```
module Booleans-syntax
exports
  sorts Bool

  context-free syntax
    "true"          -> Bool
    "false"         -> Bool
    Bool "&" Bool -> Bool {left}
    Bool "|" Bool -> Bool {right}
```

Concrete Booleans defined with visit

Using the visit construct that we have described above, we can write the definitions for the functions `&` and `|` as follows:

```
module Bool-examples1

import-syntax Booleans-syntax;

fun Bool reduce(Bool B) {
  visit bottom-up B {
    case true & <Bool B2>   => <B2>
    case false & <Bool B2> => false

    case true | true       => true
```

```
    case true | false      => true
    case false | true       => true
    case false | false      => false
  }
}
```

Observe that there are two styles:

- In the definition for `&`, we use variables on the left-hand side: the visit is needed to fully normalize the result.
- In the definition of `|`, we use a truth table.

Abstract Booleans defined with visit

In the above example we used concrete syntax for Booleans expressions. It also possible to define Booleans as abstract terms.

```
module Bool-abstract

data Bool btrue;
data Bool bfalse;
data Bool band(Bool, Bool);
data Bool bor(Bool, Bool);

fun Bool reduce(Bool B) {
  visit bottom-up B {
    acase band(btrue, <Bool B2>)  => <B2>      %% Use Variables
    acase band(bfalse, <Bool B2>) => bfalse

    acase bor(btrue, btrue)       => btrue    %% Use a truth table
    acase bor(btrue, bfalse)      => btrue
    acase bor(bfalse, btrue)      => btrue
    acase bor(bfalse, bfalse)     => bfalse
  }
}
```

First, type declarations are used to define the abstract syntax of the type Bool. Next, a similar reduce function is defined as before, but now we use abstract patterns.

Abstract Booleans defined with arules

In ASF, values are always reduced to a normal form before they are created. For some applications this normalization or canonicalization feature is very handy. We introduce the following syntax, which can also help in the transformation of old ASF+SDF programs to Rascal:

```
module Bool-rules

data Bool btrue;
data Bool bfalse;
data Bool band(Bool, Bool);
data Bool bor(Bool, Bool);

arules(Bool) {
  band(btrue, <Bool B2>)  => <B2>
  band(bfalse, <Bool B2>) => bfalse

  bor(btrue, btrue)       => btrue
  bor(btrue, bfalse)      => btrue
}
```

```
    bor(bfalse, btrue)    => btrue
    bor(bfalse, bfalse)   => bfalse
}
```

These rules are applied on every Bool that is constructed. Like in ASF+SDF it is the responsibility of the programmer to make sure the rules are confluent and terminating. The body of a rules definition has the same syntax and semantics as the switch construct, allowing backtracking, side-effects and checking of conditions.

There are some issues here:

- It should be disallowed to have private rules on public constructors; normalization is a global effect on public data-structures. On the other hand, constructors that are local to a module may have some private rules applied to them; but public rules on private constructors are disallowed too.
- We previously had other ideas about this feature. We introduced a `reduce` keyword that would call a certain named function at the construction of every Bool. The problem with that feature is that the operational semantics (in the eyes of the Rascal programmer) really looks like first some term is to be created, and then a function will be applied to normalize it. In reality though, we want to normalize the terms in an innermost fashion, and at construction time, so that the terms on the left-hand side of the rules are never actually created or allocated.

Caution

The following seems more declarative and more close to the semantics of rewrite rules being applied everywhere and in any order:

```
rule and(true,B2) => B2
rule and(true,B2) : { print("rule!"); yield B2 }
```

Concrete Booleans defined with rules

In a similar fashion, the concrete syntax version of Booleans can be defined using rules:

```
module Bool-conc-rules

import Booleans-syntax;

rules(Bool) {
    true & <Bool B2>    => <B2>
    false & <Bool B2>   => false

    true | true         => true
    true | false        => true
    false | true        => true
    false | false       => false
}
```

Abstract Graph datatype

In the Meta-Environment we use an abstract data type to exchange data representing graphs. It can be defined as follows.

```
module Graph

data Graph graph(NodeList nodes,
                 EdgeList edges,
                 AttributeList attributes);
```

```
type list[Node] Nodelist;

data Node node(NodeId id,
               AttributeList attributes);

data Node subgraph(NodeId id,
                  NodeList nodes,
                  EdgeList edges,
                  AttributeList attributes);

data NodeId id(term id);

type list[Attribute] AttributeList;

data Attribute bounding-box(Point first, Point second);
data Attribute color(Color color);
data Attribute curve-points(Polygon points);
data Attribute direction(Direction direction);
data Attribute fill-color(Color color);
data Attribute info(str key, value value);
data Attribute label(str label);
data Attribute tooltip(str tooltip);
data Attribute location(int x, int y);
data Attribute shape(Shape shape);
data Attribute size(int width, int height);
data Attribute style(Style style);
data Attribute level(str level);
data Attribute file(File file);
data Attribute file(value file);

data Color rgb(int red, int green, int blue);

data Style bold | dashed | dotted | filled | invisible | solid;

data Shape box | circle | diamond | egg | ellipse | hexagon |
           house | octagon | parallelogram | plaintext |
           trapezium | triangle;

data Direction forward | back | both | none;

type list[Edge] Edgelist;

data Edge edge(NodeId from,
               NodeId to,
               AttributeList attributes);

type list[Point] Polygon;

data Point point(int x, int y);
```

Tree traversal

Here is the binary tree example that we use in explaining traversal functions in ASF+SDF.

```
module BTree-syntax
imports basic/Integers

exports
```

```
sorts BTREE
context-free syntax
  Integer      -> BTREE
  f(BTREE,BTREE) -> BTREE
  g(BTREE,BTREE) -> BTREE
  h(BTREE,BTREE) -> BTREE
  i(BTREE,BTREE) -> BTREE
```

```
module BTree-Examples
import-syntax BTree-syntax;

%% Ex1: Count leaves in a BTREE
%% Idea: int N : T generates alle Integer leaves in the tree

fun int cnt(BTREE T) {
  return size({N | int N : T})
}

%% Ex1: an equivalent, more purist, version of the same function:
fun int cnt(BTREE T) {
  return size({N | Integer N : T})
}

%% Ex1: alternative solution using trafo functions:

fun int cnt(BTREE T) {
  int C = 0;
  visit T {
    case <Integer N> : C = C+1
  };
  return C;
}

%% Ex2: Sum all leaves in a BTREE
%% NB sum is a built-in that adds all elements in a set or list.
%% Here we see immediately the need to identify
%% - the built-in sort "int"
%% - the syntactic sort "Integer"

fun int sumtree(BTREE T) {
  return sum({N | int N : T});
}

%% Ex2: using accumulator

fun int cnt(BTREE T) {
  int C = 0;
  visit T {
    case <Integer N> : C = C+N
  };
  return C;
}

%% Ex3: Increment all leaves in a BTREE
%% Idea: using the construct "visit T { ... }" visit all leaves in
%% the tree T that match an integer and replace each N in T by N+1.
%% The expression as a whole returns the modified term.
%% Note that two conversions are needed here:
```

```
%% - from int to NAT to match subterms
%% - from int to NAT to convert the result of addition into
%%   a NAT tree

fun BTREE inc(BTREE T) {
  visit T {
    case <Integer N>: yield (N + 1)
  };
}

%% Ex4: full replacement of g by i
%% The whole repertoire of traversal functions is available:
%% - visit bottom-up T { ... }
%% - visit bottom-up-break T { ... }
%% - etc.
%% A nice touch is that these properties are not tied to the
%% declaration of a traversal function (as in ASF+SDF) but to
%% its use.

fun BTREE frepl(BTREE T) {
  visit bottom-up T {
    case g(<BTREE T1>, <BTREE T2>) =>
      i(<BTREE T1>, <BTREE T2>)
  };
}

%% Ex5: Deep replacement of g by i

fun BTREE frepl(BTREE T) {
  visit bottom-up-break T {
    case g(<BTREE T1>, <BTREE T2>) =>
      i(<BTREE T1>, <BTREE T2>)
  };
}

%% Ex6: shallow replacement of g by i (i.e. only outermost
%% g's are replaced);

fun BTREE srepl(BTREE T) {
  visit top-down-break T {
    case g(<BTREE T1>, <BTREE T2>) =>
      i(<BTREE T1>, <BTREE T2>)
  };
}

%% Ex7: We can also add the first/td directives to all generators
%% (where "all td" would be the default):

fun set[BTREE] find_outer_gs(BTREE T) {
  return
  { S | STATEMENT S : top-down-break T,
    g(<BTREE T1>, <BTREE T2>) := S };
}

%% Ex8: accumulating transformer that increments leaves with
%% amount D and counts them
fun tuple[int, BTREE] count_and_inc(BTREE T, int C, int D) {
  int C = 0;
```



```
visit T {
  case <Integer N>: { C = C + 1; N = N + D; yield [| <N> |]}
};
return <C, T>;
}
```

Substitution in Lambda

Below a definition of substitution in lambda expressions. It would be nice to get this as simple as possible since it is a model for many binding mechanisms. It is also a challenge to write a generic substitution function that only depends on the syntax of variables and argument binding.

```
module examples/Lambda/Lambda-syntax

exports
sorts Var %% variables
      Exp %% expressions
lexical syntax
  [a-z]+          -> Var
context-free syntax
  "prime" "(" Var ")" -> Var  %% generate unique name
  Var          -> Exp  %% single variable
  "fn" Var ">=>" Exp -> Exp  %% function abstraction
  Exp Exp      -> Exp  %% function application
```

Examples:

```
module Lambda-Examples

import-syntax Lambda-syntax;

fun set[Var] allVars(Exp E) {
  return {V | Var V : E}
}

fun set[Var] boundVars(Exp E) {
  return {V | fn <Var V> => <Exp E1> : E};
}

fun set[Var] freeVars(Exp E) {
  return allVars(E) - boundVars(E);
}

%% Generate a fresh variable if V does not occur in
%% given set of variables.

fun Var fresh(Var V, set[Var] S) {
  if (V in S){ return prime(V); } else {return V;};
}

%% Substitution: replace all occurrences of V in E2 by E1

fun Exp subst(Var V1, Exp E1, Exp E2) {

  switch E2 {
    case <Var V2>: if(V1 != V2){ yield [| <V2> |] }
  }
```

```

    case <Var V2>: if(V1 == V2){ yield [| <E1> |] }

    case <Exp Ea> <Exp Eb>: {
      Exp EaS = subst(V, E, Ea);
      Exp EbS = subst(V, E, Eb);
      return [| <Exp EaS> <Exp EbS> |];
    }

    case fn <Var V2> => <Var Ea>:
      if (V1 == V2) { yield [| fn <V2> =\> <Ea> |] }

    case fn <Var V2> => <Exp Ea>:
      if(V1 != V2 and not(V1 in freeVars(E2) and
        V2 in freeVars(E1))){
        Exp E1S = subst(V1, E1, Ea);
        yield [| fn <Var V2> =\> <Exp E1S> |];
      }

    case fn <Var V2> => <Exp Ea>:
      if(V1 != V2 and V1 in freeVars(Ea) and
        V2 in freeVars(E1)){
        Var V3 = fresh(V2, freeVars(Ea) + freeVars(E1));
        Exp EaS = subst(V1, E1, subst(V2, V3, E2));
        yield [| fn <Var V3> =\> <Exp EaS> |];
      }
  };
}

```

Renaming in Let

```

module Let-syntax
exports
  sorts Var %% variables
         Exp %% expressions
lexical syntax
  [a-z]+          -> Var
context-free syntax
  Var             -> Exp
  "let" Var "=" Exp "in" Exp "end" -> Exp

```

Examples:

```

module Let-Example
imports Let;

%% Rename all bound variables in an Exp
%% Version 1: purely functional
%% Exp: given expression to be renamed
%% rel[Var,Var]: renaming table
%% Int: counter to generate global variables

fun Exp rename(Exp E, rel[Var,Var] Rn, Int Cnt) {
  switch E {
    case let <Var V> = <Exp E1> in <Exp E2> end: {
      Var Y = "x" + Cnt;  %% this + operator concatenates
                        %% (after converting the int to str)
      int Cnt1 = Cnt + 1;
      Exp E1R = rename(E1, Rn, Cnt);

```

```
    Exp E2R = rename(E2, {<V, Y>} + Rn, Cnt1);
    return make let <Var Y >= <Exp E1R>
        in
            <Exp E2R>
        end;
    }

    case <Var V>: return Rn(V)

    default: return E
    };
}
```

Renaming in Let using globals

Here is the same renaming function now using two global variables.

```
module Let-Example

import-syntax Let;

%% Rename all bound variables in an Exp
%% Version 2: using global variables
%% Cnt: global counter to generate fresh variables
%% rel[Var,Var]: global renaming table

global int Cnt = 0;
global rel[Var,Var] Rn = {};

fun Var newVar() {
    global int Cnt;
    Cnt = Cnt + 1;
    return "x" + Cnt
}

fun Exp rename(Exp E) {
    global int Cnt;
    global rel[Var,Var] Rn;
    switch E {
    case let <Var V> = <Exp E1> in <Exp E2> end: {
        Var Y = newVar();
        Rn = {<V, Y>} + Rn;
        Exp E1R = rename(E1);
        Exp E2R = rename(E2);
        return [| let <Var Y >= <Exp E1R>
            in
                <Exp E2R>
            end
        |]
    }

    case <Var V>: return Rn(V)

    default: return E
    };
}
```

Concise Pico Typechecker

The following example shows the tight integration ASF with comprehensions.

```

module Typecheck

imports Pico-syntax;
imports Errors;

subtype Env rel[PICO-ID,TYPE];

fun list[Error] tcp(PROGRAM P) {
  switch P {
    case begin <DECLS Decl> <{STATEMENT ";"}* Series> end: {
      Env Env = {<Id, Type> |
        [| <PICO-ID Id> : <TYPE Type> |] : Decl};
      return [ tcst(S, Env) | Stat S : Series ]
    }
  };
}

fun list[Error] tcst(Stat Stat, Env Env) {
  switch Stat {
    case [| <PICO-ID Id> = <EXP Exp>|]: {
      TYPE Type = Env(Id);
      return type-of(Exp, Type, Env);
    }

    case if <EXP Exp> then <{STATEMENT ";"}* Stats1>
      else <{STATEMENT ";"}* Stats2> fi:
      return type-of(Exp, natural, Env) +
        tcs(Stats1, Env) + tcs(Stats2, Env)

    case while <EXP Exp> do <{STATEMENT ";"}* Stats1> od:
      yield type-of(Exp, natural, Env) + tcs(Stats, Env)
  };
}

fun list[Error] type-of(Exp E, TYPE Type, Env Env) {
  switch E {
    case <NatCon N>: if(Type == natural){ return [|; }

    case <StrCon S>: if(Type == string) { return [|; }

    case <PICO-ID Id>: {
      TYPE Type2 = Env(Id);
      if(Type2 == Type) { return [|; }
    }

    case <EXP E1> + <EXP E2>:
      if(Type == natural){
        return type-of(E1, natural, Env) +
          type-of(E2, natural, Env);
      }

    case <EXP E1> * <EXP E2>:
      if(Type == natural){
        return type-of(E1, natural, Env) +
          type-of(E2, natural, Env);
      }
  }
}

```

```
        type-of(E1, natural, Env);
    }

    case <EXP E1> || <EXP E2>:
        if(Type == string){
            return type-of(E1, string, Env) +
                type-of(E1, string, Env)
        }

        default: return [error("Incorrect type")]
    };
}
```

Pico evaluator

```
module Pico-eval
imports pico/syntax/Pico;

type int PICO-VALUE;
type str PICO-VALUE;

type map[PICO-ID, PICO-VALUE] VEnv;

fun VEnv evalProgram(PROGRAM P){
    switch P {
        case begin <DECLS DeclS> <{STATEMENT ";"}* Series> end: {
            VEnv Env = evalDecls(DeclS);
            return evalStatements(Series, Env);
        }
    }
};

fun VEnv assign(VEnv Env, PICO-ID Id, PICO-VALUE V){
    Env(Id) = V;
    return Env
}

fun PICO-VALUE valueOf(VEnv Env, PICO-ID Id){
    return Env(Id);
}

fun VEnv evalDecls(DECLS DeclS){
    VEnv Env = {};
    visit DeclS {
        case <PICO-ID Id> : string: Env = assign(Env, Id, "")
        case <PICO-ID Id> : natural: Env = assign(Env, Id, 0)
    }
}

fun VEnv evalStatements({STATEMENT ";"}* Series, VEnv Env){
    switch Series {
        case <STATEMENT Stat>; <{STATEMENT ";"}* Series2>: {
            Env Env2 = evalStatement(Stat, Env);
            return evalStatements(Series2, Env2);
        }
        case [] || []: return Env
    }
}
```

```

fun VEnv evalStatement(STATEMENT Stat, VEnv Env){
  switch Stat {
    case [| <PICO-ID Id> = <EXP Exp> |]: {
      PICO-VALUE Val = evalExp(Exp, Env);
      return assign(Env, Id, Val)
    }

    case if <EXP Exp> then <{STATEMENT ";"* Stats1}>
      else <{STATEMENT ";"* Stats1}> fi:{
      PICO-VALUE Val = evalExp(Exp, Env);
      if(Val != 0)
        return evalStatments(Stats1, Env)
      else
        return evalStatements(Stats2, Env)
    }

    case while <EXP Exp> do <{STATEMENT ";"* Stats1}> od:{
      PICO-VALUE Val = evalExp(Exp, Env);
      if(Val != 0)
        return Env
      else {
        VEnv Env2 = evalStatements(Stats1, Env);
        return evalStatement(Stat, Env2)
      }
    }
  };
};

fun PICO-VALUE evalExp(Exp exp, VEnv Env) {
  switch exp {
    case <NatCon N>: return N

    case <StrCon S>: return S

    case <PICO-ID Id>: return valueOf(Env, Id)

    case <EXP exp1> + <EXP exp2>: {
      Natural nat1 = evalExp(exp1, Env);
      Natural nat2 = evalExp(exp2, Env);
      return nat1 + nat2;
    }

    case <EXP exp1> - <EXP exp2>: {
      Natural nat1 = evalExp(exp1, Env);
      Natural nat2 = evalExp(exp2, Env);
      return nat1 - nat2;
    }

    case <EXP exp1> || <EXP exp2>: {
      StrCon str1 = evalExp(exp1, Env);
      StrCon str2 = evalExp(exp2, Env);
      return concat(str1, str2);
    }
  }
}

```

Pico evaluator with globals

Here is the same evaluator but now using a global variable to represent the value environment.

```
module Pico-eval
imports pico/syntax/Pico;

type int PICO-VALUE;
type str PICO-VALUE;

type map[PICO-ID, PICO-VALUE] VEnv;

fun void evalProgram(PROGRAM P){
  switch P {
    case begin <DECLS Decls> <{STATEMENT ";"* Series> end: {
      evalDecls(Decls);
      evalStatements(Series);
    }
  }
};

fun void assign(PICO-ID Id, PICO-VALUE V){
  global Venv Env;
  Env(Id) = V;
  return;
}

fun PICO-VALUE valueOf(PICO-ID Id){
  global Venv Env;
  return Env(Id);
}

fun VEnv evalDecls(DECLS Decls){
  global VEnv Env = {};
  visit Decls {
    case <PICO-ID Id> : string: assign(Id, "")
    case <PICO-ID Id> : natural: assign(Id, 0)
  }
}

fun void evalStatements({STATEMENT ";"* Series){
  global VEnv Env;
  switch Series {
    case <STATEMENT Stat>; <{STATEMENT ";"* Series2>: {
      evalStatement(Stat);
      evalStatements(Series2);
      return
    }
    case [] | []: return
  }
}

fun void evalStatement(STATEMENT Stat){
  global VEnv Env;
  switch Stat {
    case [] <PICO-ID Id> = <EXP Exp> []: {
      PICO-VALUE Val = evalExp(Exp);
      assign(Id, Val);
    }
  }
}
```

```

        return
    }

    case if <EXP Exp> then <{STATEMENT ";"}* Stats1>
        else <{STATEMENT ";"}* Stats1> fi:{
        PICO-VALUE Val = evalExp(Exp);
        if(Val != 0) {
            evalStatements(Stats1);
            return
        } else {
            evalStatements(Stats2);
            return
        }
    }

    case while <EXP Exp> do <{STATEMENT ";"}* Stats1> od:{
        PICO-VALUE Val = evalExp(Exp);
        if(Val != 0)
            return
        else {
            evalStatements(Stats1);
            evalStatement(Stat);
            return
        }
    }
}

};

fun PICO-VALUE evalExp(Exp exp) {
    global VEnv Env;
    switch exp {
        case <NatCon N>: return N

        case <StrCon S>: return S

        case <PICO-ID Id>: return valueOf(Id)

        case <EXP exp1> + <EXP exp2>: {
            Natural nat1 = evalExp(exp1);
            Natural nat2 = evalExp(exp2);
            return nat1 + nat2;
        }

        case <EXP exp1> - <EXP exp2>: {
            Natural nat1 = evalExp(exp1);
            Natural nat2 = evalExp(exp2);
            return nat1 - nat2;
        }

        case <EXP exp1> || <EXP exp2>: {
            StrCon str1 = evalExp(exp1);
            StrCon str2 = evalExp(exp2);
            return concat(str1, str2);
        }
    }
}

```


Pico control flow extraction

Caution

Is the definition of CP ok? Better to use a datatype?

```

module Pico-controlflow

import-syntax pico/syntax/Pico;

type EXP CP;
type STATEMENT CP;

type tuple[set[CP] entry,
           rel[CP,CP] graph,
           set[CP] exit] CFSEGMENT;

fun CFSEGMENT cflow({STATEMENT ";"* Stats}){
  switch Stats {
    case <STATEMENT Stat> ; <{STATEMENT ";"* Stats2}>: {
      tuple[set[CP] En1, rel[CP,CP] R1, set[CP] Ex1] =
        cflow(Stat);
      tuple[set[CP] En2, rel[CP,CP] R2, set[CP] Ex2] =
        cflow(Stats2);
      return <En1, R1 + R2 + product(Ex1, En2), Ex2>
    }
    case [] | []: return <{}, {}, {}>
  }
};

fun CFSEGMENT cflow(STATEMENT Stat){
  switch Stat {
    case while <EXP Exp> do <{STATEMENT ";"* Stats}> od : {
      tuple[set[CP] En, rel[CP,CP] R, set[CP] Ex] =
        cflow(Stats);
      return <{Exp}, product({Exp}, En) + R +
        product(Ex, {Exp}),
        {Exp}>;
    }
    case if <EXP Exp> then <{STATEMENT ";"* Stats1}>
      else <{STATEMENT ";"* Stats2}> fi: {
      tuple[set[CP] En1, rel[CP,CP] R1, set[CP] Ex1] =
        cflow(Stats1);
      tuple[set[CP] En2, rel[CP,CP] R2, set[CP] Ex2] =
        cflow(Stats2);
      return < {Exp},
        product({Exp}, En1) +
        product({Exp}, En2) + R1 + R2,
        Ex1 + Ex2
        >;
    }
    case <STATEMENT Stat>: return <{Stat}, {}, {Stat}>
  }
};

```

Pico use def extraction

```
module Pico-use-def

import-syntax pico/syntax/Pico;

fun rel[PICO_ID, EXP] uses(PROGRAM P) {
  return {<Id, E> | EXP E : P, PICO_ID Id := E}
}

fun rel[PICO_ID, STATEMENT] defs(PROGRAM P) {
  return {<Id, S> | STATEMENT S : P,
    [| <PICO_ID Id> := <EXP Exp> |] := S}
}
```

The above uses a "matching condition" to decompose S. The problem solved is that we want to have a name for the whole assignment *and* for the lhs identifier. Also note that, compared to older definitions of these functions, the identifier is placed as first element in each tuple.

Pico uninitialized variables

```
module Pico-uninit

import-syntax pico/syntax/Pico;

import Pico-controlflow;
import Pico-use-def;

fun set[PICO_ID] uninit(PROGRAM P) {
  rel[EXP, PICO_ID] Uses = uses(P);
  rel[PICO_ID, STATEMENT] Defs = defs(P);
  CFSEGMENT CFLOW = cflow(P);
  set[CP] Root = CFLOW.entry;
  rel[CP, CP] Pred = CFLOW.graph;

  return {Id | tuple[EXP E, PICO_ID Id] : Uses,
    E in reachX(Root, Defs[Id], Pred)
  };
}
```

Pico common subexpression elimination

```
module Pico-common-subexpression

import-syntax pico/syntax/Pico;
import Pico-controlflow;
import Pico-use-def;

fun PROGRAM cse(PROGRAM P) {
  rel[PICO_ID, STATEMENT] Defs = defs(P);
  rel[CP, CP] Pred = cflow(P).graph;
  map[EXP, PICO_ID] replacements =
    {<E2, Id> | STATEMENT S : P,
      [| <PICO_ID Id> := <EXP E> |] := S,
      Id notin E,
      EXP E2 : reachX({S}, Defs[Id], Pred)
    };
}
```

```
visit P {  
  case <EXP E>: if(PICO_ID Id := replacements(E)) yield Id  
};  
}
```

Paraphrased: Replace in P all expressions E2 by Id, such that

- P contains a statement S of the form Id := E,
- Id does not occur in E,
- E2 can be reached from S,
- There is no redefinition of Id between S and E2.

Caution

The following is to be discussed and will probably be omitted.

Also note that we could factor out the assignment pattern to make cse more generic if we introduce patterns as first class citizens:

```
fun PROGRAM cse(PROGRAM P,  
  pat STATEMENT Assign(PICO-ID Id, EXP E)) {  
  ...  
  rel[EXP, PICO-ID] replacements :=  
    {<E2, Id> | Assign S : P,  
      Id notin E,  
      EXP E2 : reachX({S}, Defs(Id), Pred)  
    };  
  ...  
}
```

Example invocations (Pico style)

```
cse(P, <PICO-ID Id> := <EXP E>)
```

or (Cobol style):

```
cse(P, move <EXP E> to <PICO-ID Id>)
```

Note that the order of variables in the pattern and its declaration may differ.

It is to be determined how the instantiation of a pattern looks, e.g.

```
Assign([|x|], [| y = 1 |])
```

Pico constant propagation

```
module Pico-constant-propagation  
  
import-syntax pico/syntax/Pico;  
import Pico-controlflow;  
import Pico-use-def;  
  
fun Boolean is_constant(EXP E) {  
  switch E {  
    case <NatCon N> => true
```

```

    case <StrCon S> => true

    case <EXP E>      => false
  }
}

fun PROGRAM cp(PROGRAM P) {
  rel[PICO_ID, STATEMENT] Defs = defs(P);
  rel[CP,CP] Pred = cflow(P).graph;

  map[PICO_ID, EXP] replacements =
    {<Id2, E> | STATEMENT S : P,
      [| <PICO_ID Id> := <EXP E> |] := S,
      is_constant(E),
      PICO_ID Id2 : reachX({S}, Defs[Id], Pred),
      Id2 == Id
    };

  visit P {
    case <PICO-ID Id>: if(EXP E := replacements[Id]) yield E
  };
}

```

Paraphrased: Replace in P all expressions Id2 by the constant E, such that

- P contains a statement S of the form Id := E,
- E is constant,
- Id2 can be reached from S,
- Id2 is equal to Id,
- There is no redefinition of Id between S and Id2.

Pico Reaching definitions

Recall the equations construct as used, for example, in the reaching definitions example in the Rscript guide. It computes the values of a set of variables until none of them changes any longer. The "solve" statement achieves the same effect.

```

module Pico-reaching-defs

type tuple[Stat theStat, Var theVar] Def;
type tuple[Stat theStat, Var theVar] Use;

fun set[Stat] predecessor(rel[Stat,Stat] P, Stat S) {
  return invert(P)(S)
}

fun set[Stat] successor(rel[Stat,Stat] P, Stat S) {
  return P(S)
}

fun rel[Stat, Def] reaching_definitions(rel[Stat,Var] DEFS,
                                       rel[Stat,Stat] PRED) {

  set[Stat] STATEMENT = carrier(PRED);

```

```
rel[Stat,Def] DEF = {<S,<S,V>> | tuple[Stat S, Var V]: DEFS};

rel[Stat,Def] KILL =
    {<S1, <S2, V>> | tuple[Stat S1, Var V] : DEFS,
                        tuple[Stat S2, V] : DEFS,
                        S1 != S2
    };

rel[Stat,Def] IN = {};
rel[Stat,Def] OUT = DEF;

solve {
    IN = {<S, D> | int S : STATEMENT,
                Stat P : predecessor(PRED,S),
                Def D : OUT[P]};
    OUT = {<S, D> | int S : STATEMENT,
                Def D : DEF[S] + (IN[S] - KILL[S])}
};
return IN;
}
```

Structured lexicals: numbers

Given the SDF definition:

```
sorts Digit Number Real
lexical syntax
[0-9]          -> Digit
Digit+         -> Number
Number "." Number -> Real
```

we can write normalization rules for Number that remove leading zeros:

```
rules(Number){
    number("0" <Digit+ Ds>) => number(<Ds>)
}
```

Note that a character class can be used instead of the sort Digit:

```
rules(Number){
    number("0" <[0-9]+ Ds>) => number(<Ds>)
}
```

A truncation function on Real can replace the mantissa by 0:

```
fun Real truncate(Real R){
    switch R {
        case real(<Number Num> "." <Digit+ Ds>) => real(<Num> "." "0")
    }
}
```

Structured Lexicals: strings

Given the SDF definition:

```
sorts String NQChar
lexical syntax
~["\"]          -> NQChar
["\"] NQChar* ["\"] -> String
```

A function that removes the *first* newline from a string can be written as:

```
fun String removeFirstNL(String S){
  switch S {
    case string("\"" <NQChar* Chars1> "\n" <NQChar* Chars2> "\"") =>
      string("\"" <Chars1> <Chars2> "\"")
  }
}
```

A function that removes all newline from a string:

```
fun String removeAllNL(String S){
  visit S {
    case string("\"" <NQChar* Chars1> "\n" <NQChar* Chars2> "\"") =>
      string("\"" <Chars1> <Chars2> "\"")
  }
}
```

Symbol table with scopes

Here is a (probably naive) implementation of a symbol table that maintains a list of numbered scopes as well as a (Name, Value) mapping in each scope. Note that we introduce parameterized modules to do this right.

```
module SymTable[&Name, &Value]

%% A scope-oriented symbol table.
%% Each scope consists of a map from names to values.
%% This is more intended to explore whether this can be expressed
%% *at all* than that the datatype is well designed.

type rel[&Name, &Value] ScopeMap;
type int ScopeId;
data STable[&Name, &Value] stable(ScopeId scope,
                                   rel[int, ScopeMap] scopes);

%% Create a new, empty, table
fun STable[&Name, &Value] new_table(){
  return stable(0, {<0, {}}>);
}

%% Create a new, non-empty, table
fun STable[&Name, &Value] new_table(ScopeId scope,
                                   rel[int, ScopeMap] scopes){
  return stable(scope, scopes);
}

%% Update, in a given scope, the value of a variable
fun STable[&Name, &Value] update(STable[&Name, &Value] ST,
                                ScopeId scope,
                                &Name N,
                                &Value V){
  ST.scopes(scope) = V;
  return new_table(scope, ST.scopes)
}

%% Get, in a given scope, the value of a variable
fun STable[&Name, &Value] value_of(STable[&Name, &Value] ST,
```

```

                                ScopeId scope,
                                &Name N){
    return ST.scopes(scope)(N)
}

%% update, in the current scope, the value of a variable
fun STable[&Name, &Value] update(STable[&Name, &Value] ST,
                                &Name N,
                                &Value V){
%%  ST.scopes(scope)(N)= V;
    return new_table(scope, ST.scopes)
}

%% Get, in the current scope, the value of a variable
fun STable value_of(STable[&Name, &Value] ST,
                    &Name N){
    return ST.scopes(ST.scope)(N)
}

%% add a new scope and make it the current scope
fun STable[&Name, &Value] new_scope(STable[&Name, &Value] ST){
    ScopeId scope = ST.scope + 1;
    return new_table(scope, ST.scopes);
}

%% switch to another scope
fun STable[&Name, &Value] switch_scope(STable[&Name, &Value] ST,
                                       ScopeId scope){
    return new_table(scope, ST.scopes);
}

```

Innerproduct

[Example taken from TXL documentation]

Define innerproduct on vectors of integers, e.g. (1 2 3).(3 2 1) => 10.

```
module examples/Vectors/Vector-syntax
```

```
exports
```

```
  imports basic/Integers
```

```
sorts Vector
```

```
context-free syntax
```

```
  "(" Integer* ")"    -> Vector
```

```
  Vector "." Vector   -> Integer
```

```
module Innerproduct
```

```
import Vector-syntax;
```

```
fun int innerProduct(Vector V1, V2){
```

```
  if ( <Integer N1> <Integer* Rest1> := V1 and
```

```
      <Integer N2> <Integer* Rest2> := V2
```

```
  )
```

```
    return (N1*N2) + innerProduct( [| (<Rest1>) |], [| (<Rest2>) |] )
```

```
  else
```

```
    return 0
}
```

Bubble sort

[Example taken from TXL documentation]

```
module Bubble

fun Integer* sort(Integer* Numbers){
  visit Numbers {
    case <Integer* Rest1>
      <Integer N1> <Integer N2>
      <Integer* Rest2>:
    if(N1 > N2){
      return sort( [| <Integer* Rest1>
                    <Integer N2> <Integer N1>
                    <Integer* Rest>
                    |] );
    }
  };
  return Numbers
}
```

Another way to write this is:

```
module Bubble

fun Integer* sort(Integer* Numbers){
  visit Numbers {
    case <Integer N1> <Integer N2>:
    if( N1 > N2 )
      yield [| <Integer N2> <Integer N1> |]
    };
  return Numbers
}
```

The visit will replace all adjacent pairs that are in the wrong order in the current list. This is repeated until no more changes are possible.

Caution

In the two above examples it is not elegant that the pattern following the case needs no quoting, and that the pattern following the return/yield has to be quoted.

Generic Bubble sort

Caution

The current syntax does not yet allow type variables in patterns. What are the implications for the implementation? Do we want this generality? Proposal: no since the ramifications are unclear.

Here is a generic bubble sort wich uses type parameters (&ELEM) and a function parameter.

```
module Bubble-Gen

fun &Elem* sort(&Elem* Elements,
               fun bool GreaterThan(&Elem, &Elem)){
```



```
visit repeat Elements {
  case <&Elem E1> <&Elem E2>:
    if(GreaterThan(E1, E2))
      yield <&Elem E2> <&Elem E1>
}
```

Of course, we can write a generic sort on arbitrary lists.

Read-Eval-Print Loop (REPR) [Needs further discussion]

For the scripting of application it is important to have a command language and read-eval-print loop. Here is an attempt. The command prompt is ">".

```
> import lang.java.syntax.Main as Java
> str source := read("program.java");
> CU program := Java.CU.parse(source);

> fun int count(CU P) {
  int cnt := 0;
  visit P {
    Java.Statements.IF => cnt++;
  }
  return cnt
}

> count(program)
17
```

There are several innovations here:

- The import associates a name with the imported module.

Caution

Work out the details.

Note

This means that "grammar" and "rule" become notions that can be manipulated.

- There is a read functions that reads a text file into a string.

Caution

We need an io library that reads/writes strings and data values. The current version is not yet ok.

- We associate a parse function with every non-terminal in a grammar.
- The notation `Java.Statements.IF` consists of three parts:
 - Language name
 - Sort name

- Rule name (currently implemented with the "cons" attribute).

It can be used as pattern. Other potential uses are as generator:

```
{S | Java.Statements.IF S : P}
```

It generates all if statements in P.

The Rascal standard library

In this section we summarize the (over 300!) functions in the Rascal Standard Library.

Main functions

The main functions of the library are listed in the following table.

Table 1. Main functions and operators in Rascal Standard Library

	bool	int	double	str	loc	list	set	map	rel	tuple_
==	x	x	x	x	x	x	x	x	x	x
<	x	x	x	x	x	x	x	x	x	x
+	x	x	x	x		x	x			x
-	x	x	x				x			
*	x	x	x							
/		x	x							
&							x	x	x	
size				x		x	x	x		x
get				x		x				x
arb	x	x	x			x	x	x	x	
toInt	x	-	x	x						
toDouble	x	x	-	x						
toString	x	x	x	x	x	x	x	x	x	x
toList				x		-	x	x	x	
toSet				x		x	-	x	x	
toMap						x	x	-	x	
toRel						x	x	x	-	
reverse				x		x				
reducer						x	x	x	x	
mapper						x	x	x	x	
min	x	x	x	x		x	x			
max	x	x	x	x		x	x			
sum						x	x			
multiply						x	x			
average						x	x			
get_location					x					
set_location					x					

Notes:

- Operator `==` implements equality on all types. It is extended for data types using structural equality.
- Operator `<` implements less-than on all types.
- Operators `+`, `-`, `*`, `/` implement these operators for most types. For structured types these operators exist in three flavours: structured value op struct, structured value op element, and element op structured value.
- `size` gives the number of elements of many types.
- `get` (corresponds to the indexing notation `R[N]`) gives the N-th element of a structure.
- `arb` generates an arbitrary element from a structured value.
- `toString` converts all values to a string representation.
- `toList`, `toSet`, `toMap`, `toRel` provide conversions between structured types. They may be implemented as just a type conversion (and not a data conversion).
- `reverse` reverses the elements of ordered structured types.
- `reducer` and `mapper` take a function as argument and traverse a structured value.

Note

Due to a name clash with the type constructor `map`, we use the name `mapper` for a function that is usually called `map`. The usual function `reduce` is therefore called `reducer`.

- `min` and `max` compute the smallest (largest) of two basic values or all elements of structured values.
- `average` computes the average value of structured values that contain integers or doubles.
- `get_location` and `set_location` are access functions for values of type `loc`.

Additional functions on lists sets, maps and relations

Additional functions on lists, sets maps and relations are listed in the following table.

Table 2. Additional functions on lists, sets, maps and relations

	list	set	map	rel	graph
first	x				
rest	x				
makeString	x				
sort	x				
id		x			
power		x			
product			x	x	
compose			x	x	
invert			x	x	
complement			x	x	
domain			x	x	
range			x	x	
carrier			x	x	
domainR			x	x	
rangeR			x	x	
carrierR			x	x	
domainX			x	x	
rangeX			x	x	
carrierX			x	x	
top				x	x
bottom				x	x
reachR				x	x
reachX				x	x

Notes:

- These functions are borrowed from the Rscript library.

Caution

closure is missing

Other functions

The remaining functions take care of annotations, input/output and communication with the global database with program facts, the Program Database (PDB).

Source code

This is a very first draft of the library. When the structure is stable we will split it in modules.

Caution

The following code is maintained externally and then copied :-() to this document, so be carefull with changes.

```
module RascalStandardLibrary
```

```
%% Here is a flat list of library functions that will
%% structured in to coherent submodules.

%% --- Comparison operators -----
%% Each type defines == and <. The following
%% functions extend them.

public fun bool <=(&T A, &T B)
    return A < B or A == B

public fun bool >(&T A, &T B)
    return B < A and not A == B

public fun bool >=(&T A, &T B)
    return B < A

public fun &T min(&T A, &T B){
    if(A < B)
        return A
    else
        return B;
}

public fun &T max(&T A, &T B){
    if(A < B)
        return B
    else
        return A;
}

%% --- Boolean (bool) -----

public fun bool ==(bool B1, bool B2)
    primitive("Bool.equal")

public fun bool <(bool B1, bool B2)
    primitive("Bool.less")

public fun bool +(bool B1, bool B2)
    primitive("Bool.add")

public fun bool -(bool B1, bool B2)
    primitive("Bool.sub")

public fun bool *(bool B1, bool B2)
    primitive("Bool.mul")

%% Arbitrary boolean: arb
public fun int arb()
    primitive("Bool.arb")

public fun str toString(bool B)
    primitive("Bool.toString")

public fun int toInt(bool B)
    primitive("Bool.toInt")

public fun double toDouble(bool B)
```

```
primitive("Bool.toDouble")

%% --- Integer (int)-----

public fun bool ==(int I1, int I2)
  primitive("Int.equal")

public fun bool <(int I1, int I2)
  primitive("Int.less")

public fun int +(int I1, int I2)
  primitive("Int.add")

public fun int -(int I1, int I2)
  primitive("-Int")

public fun int *(int I1, int I2)
  primitive("Int.mul")

public fun int /(int I1, int I2)
  throws divide_by_zero(str msg)
  primitive("Int.div")

%% Arbitrary integer: arb
public fun int arb(int bgn, int end)
  throws illegal_argument(str msg) %% if(end - bgn <= 0)
  primitive("Int.arb")

public fun double toDouble(int I)
  primitive("Int.toDouble")

public fun str toString(int I)
  primitive("toStringInt")

%% --- Double (double) -----

public fun bool ==(double D1, double D2)
  primitive("Double.equal")

public fun bool ==(double D, int I)
  return D == toDouble(I)

public fun bool ==(int I, double D)
  return toDouble(I) == D

public fun bool <(double D1, double D2)
  primitive("Double.less")

public fun bool <(double D, int I)
  return D < toDouble(I)

public fun bool <(int I, double D)
  return toDouble(I) < D

public fun double +(double D1, double D2)
  primitive("Double.add")

public fun bool +(double D, int I)
```

```
    return D + toDouble(I)

public fun bool +(int I, double D)
    return toDouble(I) + D

public fun double -(double D1, double D2)
    primitive("Double.sub")

public fun bool -(double D, int I)
    return D - toDouble(I)

public fun bool -(int I, double D)
    return toDouble(I) - D

public fun double *(double D1, double D2)
    primitive("Double.mul")

public fun bool *(double D, int I)
    return D * toDouble(I)

public fun bool *(int I, double D)
    return toDouble(I) * D

public fun double /(double D1, double D2)
    throws divide_by_zero(str msg)
    primitive("Double.div")

public fun bool /(double D, int I)
    return D / toDouble(I)

public fun bool /(int I, double D)
    return toDouble(I) / D

%% Arbitrary double: arb
public fun double arb(double bgn, double end)
    throws illegal_argument(str msg) %% if(end - bgn <= 0)
    primitive("Double.arb")

public fun int toInt(double D)
    primitive("Double.toInt")

public fun str toString(double D)
    primitive("Double.toString")

%% --- Strings (str) -----

public fun bool ==(str S1, str S2)
    primitive("String.equal")

public fun bool <(str S1, str S2)
    primitive("String.less")

public fun str +(str S1, str S2)
    primitive("String.add")

public fun int size(str S)
    primitive("String.size")
```

```
public fun int get(str S, int N)
    throws out_of_range(str msg)
    primitive("String.get")

public fun int toInt(str S)
    throws cannot_convert_to_int(str msg)
    primitive("String.toInt")

public fun int toDouble(str S)
    throws cannot_convert_to_double(str msg)
    primitive("String.toDouble")

public fun list[int] toList(str S)
    throws cannot_convert_int_to_char(str msg)
    primitive("String.toList")

public fun set[int] toSet(str S)
    primitive("String.toSet")

public fun str reverse(str S)
    primitive("String.reverse")

%% Add functions for other methods of Java's String class

%% --- Locations -----

%% The location datatype
data location location(str filename,
                      int beginline,
                      int endline,
                      int begincol,
                      int endcol,
                      int offset,
                      int length);

public fun bool ==(loc L1, loc L2)
    primitive("Location.equal")

public fun bool <(loc L1, loc L2)
    primitive("Location.less")

public fun str toString(loc L)
    primitive("Location.toString")

public fun loc get_location(&T Subject)
    throws location_missing(str msg)
    primitive("Location.get_location")

public fun &T set_location(&T Subject, loc L)
    primitive("Location.set_location")

%% --- Lists -----

public fun bool ==(list[&T] L1, list[&T] L2)
    primitive("List.equal")

public fun bool <(list[&T] L1, list[&T] L2)
    primitive("List.less")
```



```
public fun list[&T] +(list[&T] L1, list[&T] L2)
  primitive("List.add")

public fun list[&T] +(list[&T] L, &T E)
  return L + [E];

public fun list[&T] +(&T E, list[&T] L)
  return [E] + L;

%% Size of list: size
public fun int size(list[&T] L)
  primitive("List.size")

%% Get list element: get
public fun &T get(list[&T] L, int N)
  throws out_of_range(str msg)
  primitive("List.get")

%% Arbitrary element of list: arb
public fun &T arb(list[&T] L)
  primitive("List.arb")

public fun str toString(list[&T] L)
  primitive("List.toString")

public fun str makeString(list[int] L)
  throws cannot_convert_int_to_char(str msg)
  primitive("List.makeString")

public fun set[&T] toSet(list[&T] L)
  primitive("List.toSet")

public fun map[&T, &U] toMap(list[tuple[&T, &U]] L)
  throws domain_not_unique(str msg)
  primitive("List.toMap")

public fun rel[&T] toRel(list[&T] L)
  primitive("List.toRel")

public fun list[&T] reverse(list[&T] L)
  primitive("List.reverse")

public fun &T reducer(list[&T] L, fun &T F (&T,&T), &T unit){
  &T result = unit;
  for(&T E : L){
    result = F(result, E)
  };
  return result
}

public fun list[&T] mapper(list[&T] L, fun &T F (&T,&T))
  return [F(E) | &T E : L];

%% Minimum element of a list: min
public fun &T min(list[&T] L){
  &T result = arb(L);
  for(&T E : L)
```

```
    if(less(E, result))
        result = min(result, E);
    return result
}

%% Maximum element of a list: max
public fun &T max(list[&T] L){
    &T result = arb(L);
    for(&T E : L)
        if(less(result, E))
            result = max(result, E);
    return result
}

%% Add elements of a List: sum
public fun &T sum(list[&T] L, &T zero){
    return reducer(L, +, zero);
}

%% Multiply elements of a list: multiply
public fun &T multiply(set[&T] R, &T one){
    return reducer(L, *, one);
}

%% Average of elements of a list: average
public fun &T average(list[&T] L, &T zero)
    return sum(L, zero)/size(L)

%% First element of list; first
public fun &T first(list[&T] L)
    throws empty_list(str msg)
    primitive("List.first")

%% Remaining elements of list: rest
public fun &T rest(list[&T] L)
    throws empty_list(str msg)
    primitive("List.rest")

%% Sort elements of list: sort
public fun list[&T] sort(list[&T] L, fun bool less(&T, &T))
    primitive("list.sort")

%% --- Tuples -----

public fun bool ==(tuple[&T] R, tuple[&T] S)
    primitive("Tuple.equal")

%% less
public fun bool <(tuple[&T] R, tuple[&T] S)
    primitive("Tuple.less")

%% Note the use of + on types to concatenate the types
%% of tuple elements

public fun tuple[&T + &U] +(tuple[&T] R, tuple[&U] S)
    primitive("Tuple.add")

public fun tuple[&T + &U] +(tuple[&T] R, &U E){
```

```

    return R + <E>;
}

public fun tuple[&T + &U] +(&T E, tuple[&U] R){
    return <E> + R;
}

public fun int size(tuple[&T] R)
    primitive("Tuple.size")

public fun value get(tuple[&T] R, int N)
    throws out_of_range(str msg)
    primitive("Tuple.get")

public fun str toString(tuple[&T] R)
    primitive("Tuple.toString")

%% --- Sets -----

%% ==

public fun bool ==(set[&T] R, set[&T] S)
    primitive("Set.equal")
%% <
public fun bool <(set[&T] R, set[&T] S)
    primitive("Set.less")

%% Add two sets (set union): operator +
public fun set[&T] +(set[&T] R, set[&T] S)
    primitive("Set.add")

public fun set[&T] +(set[&T] S, &T E)
    return S + {E}

public fun set[&T] +(&T E, set[&T] S)
    return {E} + S

%% Subtract two sets (set difference): operator -
public fun set[&T] -(set[&T] R, set[&T] S)
    primitive("Set.sub")

public fun set[&T] -(set[&T] S, &T E)
    return S - {E};

public fun set[&T] -(&T E, set[&T] S)
    return {E} - S;

%% Intersection of two sets: operator &
public fun set[&T] &(set[&T] R, set[&T] S)
    primitive("Set.intersection")

public fun set[&T] &(set[&T] S, &T E)
    return S * {E}

public fun set[&T] &(&T E, set[&T] S)
    return {E} * S;

%% Size of set: size

```

```
public fun int size(set[&T] R)
  primitive("Set.size")

%% Arbitrary element of set: arb
public fun &T arb(set[&T] S)
  primitive("Set.arb")

public fun str toString(set[&T] R)
  primitive("Set.toString")

public fun list[&T] toList(set[&T] S)
  primitive("List.toList")

public fun map[&T, &U] toMap(set[tuple[&T, &U]] S)
  throws non_unique_domain(str msg)
  primitive("Set.toMap")

public fun rel[&T] toRel(set[&T] S)
  primitive("Set.toRel")

public fun &T reducer(set[&T] S, fun &T F (&T,&T), &T unit){
  &T result = unit;
  for(&T E : S){
    result = F(result, E)
  };
  return result
}

public fun set[&T] mapper(set[&T] S, fun &T F (&T,&T))
  return {F(E) | &T E : S};

%% Minimum of a set: min
public fun &T min(set[&T] S){
  &T result = arb(S);
  for(&T E : S)
    result = min(result, E);
  return result
}

%% Maximum of a set: max
public fun &T max(set[&T] R){
  &T result = arb(R);
  for(&T E : R)
    result = max(result, E);
  return result
}

%% Sum elements of a Set: sum
public fun &T sum(set[&T] S, &T zero){
  return reducer(S, +, zero);
}

%% Multiply elements of a Set: multiply
public fun &T multiply(set[&T] S, &T one){
  return reducer(S, *, one);
}

%% Average of elements of a set; average
```

```
public fun &T average(set[&T] S, &T zero)
    return sum(S, zero)/size(R)

%% Powerset: power0
public fun set[set[&T]] power0(set[&T] R)
    throw unimplemented("power0")

%% Powerset: power1
public fun set[set[&T]] power1(set[&T] R)
    throw unimplemented("power0")

%% --- Maps -----

public fun bool ==(map[&T, &U] M1, map[&T, &U] M2)
    primitive("Map.equal")

public fun bool <(map[&T, &U] M1, map[&T, &U] M2)
    primitive("Map.less")

public fun map[&T] +(map[&T] R, map[&T] S)
    primitive("Map.add")

public fun map[&T] +(map[&T] S, &T E)
    return S + {E};

public fun map[&T] +(&T E, map[&T] S)
    return {E} + S

%% Subtract two maps (set difference): operator -
public fun map[&T] -(map[&T] R, map[&T] S)
    primitive("Map.sub")

public fun map[&T] -(map[&T] S, &T E)
    return S - {E}

public fun map[&T] -(&T E, map[&T] S)
    return {E} - S

%% Intersection of two maps: operator &
public fun map[&T] &(amp;T R, map[&T] S)
    primitive("Map.mul")

public fun map[&T] &(amp;T S, &T E)
    return S * {E}

public fun map[&T] &(&T E, map[&T] S)
    return {E} * S

public fun int size(map[&T] M)
    primitive("Map.size")

public fun &T arb(map[&T] M)
    primitive("Map.arb")

public fun str toString(map[&T] R)
    primitive("Map.toString")

public fun list[tuple[&T, &U]] toList(map[&T, &U] M)
```

```

    primitive("Map.toList")

public fun rel[tuple[&T, &U]] toSet(map[&T, &U] M)
    primitive("Map.toSet")

public fun rel[tuple[&T, &U]] toRel(map[&T, &U] M)
    primitive("Map.toRel")

%% Is this meaningful?
public fun &T reducer(map[&T] S, fun &T F (&T,&T), &T unit){
    &T result = unit;
    for(&T E : S){
        result = F(result, E)
    };
    return result
}

public fun map[&T] mapper(map[&T] M, fun &T F (&T,&T))
    return {F(E) | &T E : M};

%% --- Relations -----

public fun bool ==(rel[&T] R, rel[&T] S)
    primitive("Rel.equal")

public fun bool <(rel[&T] R, rel[&T] S)
    primitive("Rel.less")

%% Add two relations (union): operator +
public fun rel[&T] +(rel[&T] R1, rel[&T] R2)
    primitive("Rel.add")

public fun rel[&T] +(&T E, rel[&T] R)
    return toRel({E}) + R

public fun rel[&T] +(rel[&T] R, &T E)
    return R + toRel({E})

%% Subtract two relations (difference): operator -
public fun rel[&T] -(rel[&T] R1, rel[&T] R2)
    primitive("Rel.sub")

%% Intersection of two relations: operator &
public fun rel[&T] *(rel[&T] R1, rel[&T] R1)
    primitive("Rel.mull")

%% Size of relation: size
public fun int size(rel[&T] R)
    primitive("Rel.size")

public fun &T arb(rel[&T] R)
    primitive("Rel.arb")

public fun str toString(rel[&T] R)
    primitive("Rel.toString")

%% Note: in rel[&T], the type variable &T refers
%% to the tuple type of the relation.

```

```

public fun list[&T] toList(rel[&T] R)
  primitive("Rel.toList")

public fun set[&T] toSet(rel[&T] R)
  primitive("Rel.toSet")

public fun rel[&T] toRel(set[&T] S)
  primitive("Set.toRel")

public fun rel[&T] mapper(rel[&T] R, fun &T F (&T,&T))
  return {F(E) | &T E : R};

%% Cartesian product of two sets: product (was: x)
public fun rel[&T1, &T2] product(rel[&T1] R, rel[&T2] S)
  return {<X, Y> | &T1 X : R, &T2 Y : S}

%% Compose two relations: compose (was: o)
public fun rel[&T1, &T3] compose(rel[&T1, &T2] R,
                                rel[&T2, &T3] S)
  return {<X, Z> | tuple[&T1 X, &T2 Y1]: R,
                tuple[&T2 Y2, &T3 Z]: S, Y1 == Y2}

%% Identity relation: id
public fun rel[&T, &T] id(set[&T] S)
  return { <X, X> | &T X : S}

%% Inverse relation: invert (was: inv)
public fun rel[&T2, &T1] invert (rel[&T1, &T2] R)
  return { <Y, X> | tuple[&T1 X, &T2 Y] : R }

%% Complement of relation: complement (was: compl)
public fun rel[&T1, &T2] complement(rel[&T1, &T2] R)
  return product(domain(R), range(R)) - R

%% Domain of relation: domain
public fun set[&T1] domain (rel[&T1,&T2] R)
  return { X | tuple[&T1 X, &T2 Y] : R }

%% Range of relation: range
public fun set[&T1] range (rel[&T1,&T2] R)
  return { Y | tuple[&T1 X, &T2 Y] : R }

%% Carrier of relation: carrier
public fun set[&T] carrier (rel[&T,&T] R)
  return domain(R) + range(R)

%% Domain Restriction of a Relation: domainR
public fun rel[&T1,&T2] domainR (rel[&T1,&T2] R, set[&T1] S)
  return { <X, Y> | tuple[&T1 X, &T2 Y] : R, X in S }

%% Range Restriction of a Relation: rangeR
public fun rel[&T1,&T2] rangeR (rel[&T1,&T2] R, set[&T2] S)
  return { <X, Y> | tuple[&T1 X, &T2 Y] : R, Y in S }

%% Carrier Restriction of a Relation: carrierR
public fun rel[&T,&T] carrierR (rel[&T,&T] R, set[&T] S)
  return { <X, Y> | tuple[&T X, &T Y] : R, X in S, Y in S }

```

```

%% Domain Exclusion of a Relation: domainX
public fun rel[&T1,&T2] domainX (rel[&T1,&T2] R, set[&T1] S)
  return { <X, Y> | tuple[&T1 X, &T2 Y] : R, X notin S }

%% Range Exclusion of a Relation: rangeX
public fun rel[&T1,&T2] rangeX (rel[&T1,&T2] R, set[&T2] S)
  { <X, Y> | tuple[&T1 X, &T2 Y] : R, Y notin S }

%% Carrier Exclusion of a Relation: carrierX
public fun rel[&T,&T] carrierX (rel[&T,&T] R, set[&T] S)
  return { <X, Y> | tuple[&T1 X, &T2 Y] : R,
            X notin S, Y notin S }

%% Relations viewed as graphs

type rel[&T,&T] graph[&T];

%% Top of a Graph: top
public fun set[&T] top(graph[&T] G)
  return domain(G) - range(G)

%% Bottom of a Graph: bottom
public fun set[&T] bottom(graph[&T] G)
  return range(G) - domain(G)

%% Reachability with Restriction: reachR
public fun set[&T] reachR(set[&T] Start, set[&T] Restr,
                        graph[&T] G)
  return range(compose(domainR(G, Start),
                      carrierR(G, Restr)+))

%% Reachability with Exclusion: reachX
public fun set[&T] reachX(set[&T] Start, set[&T] Excl,
                        graph[&T] G)
  return range(compose(domainR(G, Start),
                      carrierX(G, Excl)+))

%% Shortest path between pair of nodes
public fun list[&T] shortestPathPair(&T From, &T To, graph[&T] G)
  primitive("Graph.shortestPathPair")

%% Shortest path between one node and all others
public fun set[list[&T]] shortestPathFrom(&T From, graph[&T] G)
  primitive("Graph.shortestPathFrom")

%% Shortest path between all nodes
public fun set[list[&T]] shortestPathAll(graph[&T] G)
  primitive("Graph.shortestPathAll")

%% TO DO

public fun rel[&T, &T] closure(rel[&T, &T])
  primitive("Rel.closure")

%% --- Annotations -----

%% Test whether a named annotation exists.

```



```
public fun bool has_annotation(&T Subject)
    primitive("Annotation.has_annotation");

%% Get the value of a named annotation
public fun value get_annotation(&T Subject, str Name)
    throws missing_annotation(str msg)
    primitive("Annotation.get_annotation")

%% Get all annotations
public fun map[str,value] get_annotations(&T Subject)
    primitive("Annotation.get_annotation")

%% Set the value of a named annotation
public fun &T set_annotation(&T Subject,
                             str Name, value AValue)
    primitive("Annotation.set_annotation")

%% Set all annotations
public fun &T set_annotations(&T Subject,
                              map[str, value] Annos)
    primitive("Annotation.set_annotations")

%% --- Parsing/Unparsing -----

public fun tree parseFile(str filename)
    throws file_does_not_exist(str msg)
    primitive("Parse.parseFile")

public fun tree parseString(str source)
    primitive("Parse.parseString")

public fun str unparseToString(tree Subject)
    primitive("Parse.unparseToString")

public fun str unparseToFile(tree Subject, str filename)
    throws cannot_create(str msg)
    throws write_error(str msg)
    primitive("Parse.unparseToFile")

%% --- io -----

public fun str readFile(str filename)
    throws does_not_exist(str msg)
    throws read_error(str msg)
    primitive("IO.read")

public fun &T readTerm(str filename)
    throws does_not_exist(str msg)
    throws read_error(str msg)
    throws term_error(str msg, loc l)
    primitive("IO.readTerm")

public fun void write(str filename, &T Subject)
    throws cannot_create(str msg)
    throws write_error(str msg)
    primitive("IO.write")
```

```
public fun void print(&T V)
    primitive("IO.print")

public fun void println(&T V)
    primitive("IO.println")

%% --- Interface with the Program Database -----
%% The following function provide a bare minimum and will
%% have to compared with the current PDB interface.
%% We assume one active PDB that can be opened and closed.
%% Values can be written to and read from the PDB.

%% Idea: it would be nice to model the PDB as a value
%% of type map[str, value] and to access it that way!

public fun void openPDB(str name)
    throws cannot_open(str msg)
    primitive("PDB.open")

public fun void closePDB()
    throws cannot_close(str msg)
    primitive("PDB.close")

public fun void writePDB(str name, &T val)
    throws cannot_write(str msg)
    primitive("PDB.write")

public fun &T readPDB(str name)
    throws cannot_read(str msg)
    primitive("PDB.readPDB")

%% If name is of type set[&T], then the set incr
%% is added to it. Similar for a rel[&T]

public fun void addSetPDB(str name, set[&T] incr)
    throws does_not_exist(str msg)
    primitive("PDB.addSet")

public fun void addRelPDB(str name, rel[&T] incr)
    throws does_not_exist(str msg)
    primitive("PDB.addRel")
```

Syntax Definition

See separate SDF definition.

Prototyping/implementation of Rascal

Every prototype will have to address the following issues:

- Parsing/typechecking/evaluating Rascal.
- How to implement the relational operations.
- How to implement matching.
- How to implement replacement.

- How to implement traversals.

The following options should be considered:

- Implementation of a typechecker in ASF+SDF:
 - Gives good insight in the type system and is comparable in complexity to the Rscript typechecker.
 - Work: 2 weeks
- Implementation of an evaluator in ASF+SDF.
 - Requires reimplementing of matching & rewriting in ASF+SDF.
 - Bound to be very slow.
 - Effort: 4 weeks
- Implementation of a typechecker in Rascal.
 - Interesting exercise to assess Rascal.
 - Not so easy to do without working Rascal implementation.
 - Not so easy when Rascal is still in flux.
 - Effort: 1 week
- Implementation of an evaluator in Rascal.
 - Ditto.
- Extending the current ASF+SDF interpreter.
 - This is a viable option. It requires extensions of UPTR.
 - Effort: 4 weeks
- Translation of Rascal to ASF+SDF in ASF+SDF.
 - Unclear whether this has longer term merit.
 - Allows easy experimentation and reuse of current ASF+SDF implementation.
 - Effort: 4 weeks
- Implementation of an interpreter in Java.
 - A future proof and efficient solution.
 - Requires reimplementing of matching & rewriting in Java.
 - Effort: 8 weeks.

Rascal implementation ideas

Rascal needs to support both a scripting experience as an optimized compiled language experience. Also, it needs to integrate fully with Meta-Environment and Eclipse IMP. Therefore, we have both a simple and unoptimized interpreter in mind, as well as a compiler that aggressively, but correctly, optimizes Rascal programs. The run-time of compiled programs and the interpreter will share the implementation of data-structures.

Data structures

Both the compiled code and the interpreter will run on the same data-structures which are defined by the IMP PDB project.

- We could start with the simple implementation that is now in IMP already which is based on the Java library and use the clone method to implement immutability. This will prove to be slow, but its an easy start.
- Integration with the ATerms; extend the ATerm library with all the features of the PDB, such that it becomes an implementation of the PDB's interfaces.
 - PDB's terms are typed, while ATerms are not.
 - ATerms demand canonicalization/sharing, which may prove to be hard to implement for maps, sets and relations.
 - PDB does not yet have any story for serialization.
 - ATerms will need to "implement" the PDB's interfaces which will add a dependency and seriously break other peoples code if we are not careful
 - ATerms need to be typed in order to implement correct visiting behavior when AFun's are overloaded.
- The C story is harder
 - Extension of C ATerms is hard due to the nature of C, the ATerm garbage collector, the ATerm header implementation and the amount of users of the ATerm library
- It may be a good idea to generate Rascal data-types from SDF definitions as an intermediate step, however, Rascal should still implement special code for UPTR trees for performance reasons (unlike Apigen which does not know anything about UPTR).
- The current PDB implementation does type checking at run-time. After implementing a type-checker for Rascal, we can easily add an implementation which does not do type checking at run-time in order to improve performance.
- The immutability feature of Rascal data is implemented in the data-structures and not by the compiler or the interpreter.

Interpreter

We just enumerate the thoughts that pop up once in a while:

- Write the interpreter in Java, and use it later to bootstrap the compiler which will be written in Rascal.
- Provide a REPL prompt such that experimenting can be done on-the-fly, both on the commandline, and in an Eclipse view.
- "fail" can be implemented using a Java exception, the catch will be at the choice points (switch).
- "return" can also be implemented using a Java exception; remember return can jump out of the context of a visitor that could be nested deeply in the structure of a term or a tree.
- List matching, and especially the kind of backtracking it requires will be implemented using exceptions instead of using continuations.

- Pattern matching needs to be implemented separately for both builtin data-types, abstract data types and concrete parse trees. Possibly using three "adapters" we can factor out the algorithm.
- We use apigen to bootstrap the interpreter. The interpreter will traverse the apigen object trees to implement it's functionality using separate classes.
- When the compiler is finished and bootstrapped on the interpreter, it may be worthwhile to reimplement/bootstrap the interpreter on the compiler again.

Compiler

The compiler will mainly follow the design of Mark's ASF+SDF compiler, which has proven to generate the fastest code in the world for these kinds of applications. Furthermore, these ideas have popped up:

- Bootstrap the compiler using the interpreter.
- Generate as readable function names as possible, mainly taking hints from the Rascal programs and of course from the SDF definitions.
- Generate Java code, one class per module.
- There is an issue with the globality of rewrite rules, they probably need to be collected and merged into a single factory per application. Rules apparently break modular compilation, especially if you want to optimize matching automaton
- For visitors we could first generate a tree node type reachability graph, and use it to generate a full traversal for a certain visitor. The generated visitor would not recurse into subtrees that will not be visited.
- After generating the visitors, non-recursive visits (i.e. the backbone of the grammar) can be inlined as much as possible to prevent using the stack for visiting trees.
- Inlining in general should be done very aggressively. This will allow other kinds of optimizations, like preventing superfluous condition checking. The ASF+SDF compiler does not do this yet, and it could mean a serious performance improvement. The cost is compilation time obviously, since the Java compiler is going to have to compile a lot more code.
- The simple control flow constructs of Rascal almost map one-to-one to Java
- "fail" is always in the current context/frame, so we need no exception implementation for fail.
- Like in Mark's compiler, list matching is to be implemented using nested while loops.
- "return" can be mapped to normal return statement in Java, except in the context of a visitor, where it should be an exception that is caught by the containing function of the visit, which immediately returns the result in the catch block that surrounds the call to the generated function that implements the visitor.
- Important optimizations:
 - Matching automaton:
 - Sharing prefixes (note that we can not reorder cases of a switch, or the rules?!?)
 - Common subexpression elimination
 - Constant detection and propagation
 - Aggressive inlining (where to stop?)

- Specialization and instantiation of visitors using grammars and data-type definitions
- Rascal functions with a Java body need the following:
 - Generate for each argument an Java argument with appropriate Java type.
 - For each return statement, check the type of the resulting value against the return type in the function header.
 - Catch any exceptions raised by the Java code, convert them to string and rethrow as Rascal exception.

Issues

- Which comment convention will we use? *Let's use Java style comments.*
- Extensibility of functions. *Proposal: let's not provide facilities for this.*
- How do we identify built-in sorts (bool, int, etc) with their syntactic counterparts?
- In a list comprehension: do list values splice into the list result?
- Ditto for set comprehensions.
- Unexplored idea: add (possibly lazy) generators for all types; this allows to generate, for instance, all statements in a program.
- Shopping list of ideas in Tom:
 - Named patterns to avoid building a term, i.e. `w@[[] while $Exp do $stat od []]`.
 - Anonymous variables a la Prolog: `[] while $_ do $stat od []]`.
 - String matching in patterns.
 - Tom uses the notation `%[...]%` for quoted strings with embedded `@...@` constructs that are evaluated. It also has a backquote construct.
- Shopping list of ideas from TXL:
 - "redefine" allows modification of an imported grammar.
 - An "any" sort that matches anything.

Graveyard

Don't read the following sections; they are leftovers from earlier versions of this document but may still contain material that can be reused.

Mapping features to datatypes

Note

This section has played a role during initial design; it is now outdated.
Emphasized cells indicate a new datatype/feature combination that needs to be thought out.

Table 3. Features vs datatypes

Which features work on which datatypes?	CF syntax trees	CF syntax lists	Lexical syntax trees	Lexical syntax lists	Lists	Sets	Relations	Tuples
Pattern matching	Y (CS)	Y, CS, LM	Y, PS	Y, PS, LM	Y, HT	Y, HT	Y, HT	Y
Pattern construction	Y, CS	Y, PS	Y, PS	Y	Y, HT	Y, HT	Y, HT	Y
Generator/Comprehension	N	N	N	N	LC	SC	SC	N
Complete Functions	Y	Y	Y	Y	Y	Y	Y	Y
Equations	Y, BC	Y, BC	Y, BC	Y, BC	Y	Y	Y	Y
Polymorphism	N	N	N	N	Y	Y	Y	Y
Serialization	UPTR	UPTR	UPTR	UPTR	Y	Y	Y	Y
Traversal Functions	Y	Y	Y	Y	Y	Y	Y	Y
Subtyping	N	N except character class inclusion	N,	N	Y	Y	Y	Y

- BC = Backward Compatible with ASF
- CS = Concrete Syntax
- HT= head/tail matching
- LC = List comprehension
- LM = List Matching
- N = No
- PS = Prefic Syntax
- SC = Set Comprehension
- Y = Yes

Visitor definitions (UNDECIDED and INCOMPLETE)

Visitor definitions are a new idea that borrow the programmability of Systems S's single level traversals and add them to Rascal. The idea is to be able to define the strategy annotations of visit statements and generators using a simple expression language. A definition takes as formal argument the code block of the visit statement (s), which is what needs to be done at every node (the visitor).

```

%% first recurse to the arguments, then try v,
%% which if it fails returns the original structure.
visitor bottom-up(v) = all(bottom-up(v)) ; (v <+ id)

visitor innermost(t,v) = all(innermost(t,v)) ; repeat(v <+ id)

```

We demand that all visitors are infallible, which means that when the `v` block fails, they must return a default result of the correct type. In most cases, this would be the identity (`id`).

We can also try to give these definitions a more imperative look, as if they are patterns for generating code for the visitors, as in:

```
visitor bottom-up(v) {
  all {
    bottom-up(v);
  }

  try {                                %% try is the '<+' of System S,
                                     %% only if v fails the catch block is executed.
    v;
  } catch fail(t) {
    yield t;
  }
}

%% innermost goes down and only returns after
%% nothing changes anymore

visitor innermost(v) {
  all {
    innermost(v);    %% apply this to all children first.
  }
  while (true) { {    %% then until kingdom come, apply this block:
    try {
      v;              %% if v succeeds, it has a yield or a return
                      %% statement that updates the current node.
    } catch fail(t) { %% if v fails after all, we obtain a reference
                      %% to the current node visited which we can
                      %% return;
      yield t;
    }
  }
}

visitor bottom-up-dbg(v) {
  all {
    bottom-up(v);
  }

  try {                                %% try is the '<+' of System S,
                                     %% only if v fails the catch block is executed.
    v;
  } catch fail(t) {
    printf("DBG: bottom-up visitor failed on: " + t);
    yield t;
  }
}
```

After such definitions, most of which would be in the standard library of Rascal, we can use them to program actual visits:

```
visit bottom-up (t) {
  pattern => pattern
  pattern2 : { effect; }
}
```


Outdated examples

Generating Graph files in Dot format

This example illustrates the use of a comprehension on the right-hand side of an equation.

```
module Dot-generation

imports Dot-syntax

fun Dot gen-dot(rel[ID, ID] Rel) {
  [| digraph example {
    $([ node(Tup) | <ID,ID> Tup : Rel ])
  }
  |]
}

fun DotElem* gen-node(<ID Id1, ID Id2>) {
  [| node $Id1; node $Id2; $Id1 -> $Id2 |]}
}
```

Integration with Tscripts (Outdated)

Note

It is not yet clear whether we will also include Tscript in the integration effort. For the time being, this section is considered outdated.

Introduction

It is possible to speculate on an even further integration of formalisms and combining the above amalgam of ASF+SDF and Rscript with Tscripts.

Requirements

- R12: The resulting language uses a single type system. This means that relational types (possible including syntactic objects) can be used in Tscripts.
- R13: The current "expressions" in Tscript (terms that occur at the rhs of an assignment) are replaced by calls to ASF+SDF or Rscript functions.
- R14: There is minimal duplication in functionality between ASF+SDF/Rscript/Tscript.

Different styles of Type Declarations

We have at the moment, unfortunately, a proliferation of declaration styles for types.

Functions are declared in ASF+SDF as:

```
typecheck(PROGRAM) -> Boolean
```

Observe that only the type of the parameter is given but that it does not have a name.

In Rscript we have:

```
int sum(set[int] SI) = ...
```

while in Tscript processes are declared as

```
process mkWave(N : int) is ...
```

For variables a similar story applies. Variables are declared in ASF+SDF as:

```
"X" [0-9]+ -> INT
```

In Rscript we have:

```
int X  
int X : S (in comprehensions)
```

and in Tscript we have:

```
X : int
```

In order to unify these styles, we might do the following:

- The type of an entity is always written before the entity itself.
- Formal parameters have a name.

In essence, this amounts to using the declaration style as used in Rscript. So we get:

```
Boolean typecheck(PROGRAM P) is ...  
process mkWave(int N) is ...
```

Or do we want things like:

```
function typecheck(P : PROGRAM) -> Boolean is ...  
var X -> int  
process mkWave(int N) is ...
```

Advantages are:

- The category of the entity is immediately clear (function, var, process, tool, ...).
- It is readable to further qualify the category, i.e., traversal function, hidden var, restartable tool)

Global Flow of Control

We have to settle the possible flow of control between the three entities ASF+SDF, Rscript and Tscript. Since Tscript imposes the notion of an atomic action it would be problematic to have completely unrestricted flow of control. Therefore it is logical to use Tscript for the top-level control and to limit the use of ASF+SDF and Rscript to computations within atomic actions. There is no reason to restrict the flow of control between ASF+SDF and Rscript.

What are the consequences of the above choice? Let's analyze two cases:

- Parse a file from within an ASF+SDF specification. This (and similar built-ins) that use the operating system are removed from ASF+SDF. Their effect has to be achieved at the Tscript-level which is the natural place for such primitives.
- Describe I/O for a defined language. Consider Pico extended with a read statement. Here the situation is more complicated. We cannot argue that the flow of control in the Pico program (as determined by an interpreter written in ASF+SDF) should be moved to the Tscript level since Tscript

simply does not have the primitives to express this. On the other hand, we have to interrupt the flow of control of the Pico interpreter when we need to execute a read statement. The obvious way to achieve this is

- At the Tscript level, a loop repeatedly calls the Pico interpreter until it is done.
- After each call the Pico interpreter returns with either:
 - An indication that the execution is complete (and possibly a final state and/or final value).
 - An indication that an external action has to be executed, for instance the read statement. This indication should also contain the intermediate state of the interpreter. When the external action has been executed, the Pico interpreter can be restarted with as arguments the value of the external action and the intermediate state.

Experimentation will have to show whether such a framework is acceptable.

Modularization

Rscript and Tscript have no, respectively, very limited mechanisms for modularization. ASF+SDF, however, provides a module mechanism with imports, hiding, parameters and renaming. This mechanism was originally included in ASF, was taken over by SDF and is now reused in ASF+SDF. Currently, there are not yet sufficiently large Rscripts to feel the need for modules. In Tscript, there is a strong need for restricted name spaces and for imposing limitations on name visibility in order to limit the possible interactions of a process with its surroundings and to make it possible to create nested process definitions. What are the design options we have to explore?

First, we can design a new module system that is more suited for our current requirements. The advantage is that we can create an optimal solution, the disadvantage is that there are high costs involved regarding implementation effort and migrations of existing ASF+SDF specifications to the new module scheme.

Second, we can design an add-on to the ASF+SDF module system that addresses our current needs.

Third, we can try to reuse the current ASF+SDF module system.

As a general note, parameterized modules, polymorphic types and renamings are competing features. We should understand what we want. It is likely that we do not need all of them.

Before delving into one of the above alternative approaches, let's list our requirements first.

- We need grammar modules that allow the following operations: import, renaming, deletion (currently not supported but important to have a fixed base grammar on many variations on it). Parameterization and export/hidden: unclear.
- We need function modules (ASF+SDF and Rscript) that provide: import, maybe parametrization, and export/hidden.
- We need process modules (Tscript) that provide: import, export, hiding.