

A Concise Guide to Clean StdEnv

Peter Achten

Radboud University Nijmegen, Netherlands
version august 26 2011
P.Achten@cs.ru.nl

Abstract. These notes serve as a brief guide to the Clean StdEnv, the set of library modules that come with the standard distribution. It is assumed that the reader has basic knowledge of programming in Clean.

1 Roadmap

Here is a brief overview of all Clean StdEnv modules in alphabetic order and in which section you can find its explanation.

module	section	page	module	section	page
StdArray	6	20	StdBool	4.1	7
StdChar	4.4	7	StdCharList	7	22
StdClass	2.2	3	StdEnum	5.1	14
StdFile	4.5	8	StdFunc	9	23
StdInt	4.2	7	StdList	5, 5.4	14, 17
StdMisc	10	25	StdOrdList	5.5	19
StdOverloaded	2.1	1	StdReal	4.3	7
StdString	7	22	StdTuple	8	23

2 An overloaded API

In order to appreciate the Clean StdEnv, we first need to have a look at the role of overloading. Two modules play a key role: StdOverloaded and StdClass.

2.1 StdOverloaded

This module defines all standard overloaded functions and operators. For historic reasons, the keyword used for overloading is *class*, which must not be confused with object oriented classes! The name is really an abbreviation for *type class*, or, even longer, *type constructor class*. In StdOverloaded you find the usual suspects, such as arithmetic and comparison operations.

Clean does no automatic *coercion* of values: if an `Int` value is expected, then an `Int` value should be provided, and not a `Real` or a `Bool` (even if you can think of sensible coercions). As an example, `1 + 2.5` does not type check in Clean. These kinds of coercions need to be done explicitly by you, and for this purpose `StdOverloaded` provides two families of coercion functions, viz.:

```
class toInt    a :: !a    -> Int    // Convert into Int
class toChar  a :: !a    -> Char   // Convert into Char
class toBool  a :: !a    -> Bool   // Convert into Bool
class toReal  a :: !a    -> Real   // Convert into Real
class toString a :: !a    -> {#Char} // Convert into String
```

for coercing a value of some type to the given basic type, and

```
class fromInt  a :: !Int  -> a      // Convert from Int
class fromChar a :: !Char -> a      // Convert from Char
class fromBool a :: !Bool -> a      // Convert from Bool
class fromReal a :: !Real -> a      // Convert from Real
class fromString a :: !{#Char} -> a // Convert from String
```

for coercing a basic type to a desired type. The earlier example can be transformed to the following variants:

- `toReal 1 + 2.5` (which yields 3.5 of type `Real`)
- `fromInt 1 + 2.5` (which yields 3.5 of type `Real`)
- `1 + toInt 2.5` (which yields 3 of type `Int`)
- `1 + fromReal 2.5` (which yields 3 of type `Int`)

The overloaded constants `zero` and `one` are usually used in combination with the arithmetic operators (`+`, `-`, `*`, and `/`). The `zero` and `one` instances should adhere to the usual algebraic laws that you know from high school:

$$\begin{aligned} \mathbf{zero} + x &= x = x + \mathbf{zero} \\ x &= x - \mathbf{zero} \\ \mathbf{one} * x &= x = x * \mathbf{one} \\ x &= x / \mathbf{one} \end{aligned}$$

Clean cannot enforce this, so you should make certain that these basic properties hold for your custom defined instances.

Finally, one overloaded function is defined differently from all others:

```
class length m :: !(m a) -> Int
```

This is an example of a *type constructor class*. It says that `length` is overloaded for a *type constructor* `m` that can hold values of type `a`. `StdEnv` contains only *one* instance of this overloaded function: in `StdList` the type constructor is the list type (`[]`), and the instance computes the length of a list (see also section 5.4).

2.2 StdClass

When you are developing overloading functions (i.e. functions that use overloaded operators and functions, and hence become overloaded themselves), you quickly notice that these functions use similar groups of overloaded functions. A typical example is `+`, `-`, `zero`. A function that uses these operators would normally have the following signature:

```
my_overloaded_function :: ... a ... | +, -, zero a
```

Instead of enumerating every single overloaded function or operator, you can use:

```
my_overloaded_function :: ... a ... | PlusMin a
```

Here, `PlusMin` is not an overloaded function, but a *collection* of overloaded functions or operators, and is defined in `StdClass`:

```
class PlusMin a | +, -, zero a
```

Such a group of overloaded functions can be used in the overloaded context restriction of a function. It automatically expands to the single member functions. For this reason, the same keyword *class* is used.

Another frequently occurring group of overloaded functions is `*`, `/`, `one`. For this, another group is defined in `StdClass`:

```
class MultDiv a | *, /, one a
```

(Type constructor) classes can be combined to form larger classes. An example is the following, that you can also find in `StdClass`:

```
class Arith a | PlusMin, MultDiv, abs, sign, ~ a
```

An overloaded function that uses (a subset of) these overloaded functions can have signature:

```
my_overloaded_function :: ... a ... | Arith a
```

which is much shorter than:

```
my_overloaded_function :: ... a ... | +, -, zero, *, /, one, abs, sign, ~ a
```

The classes `IncDec` and `Enum` are used to create lists with the *dot-dot* notation (see also section 5.1).

The class `Eq` contains the overloaded equality test operator `==`, and uses it to add the inequality test `<>`, which is defined in terms of `==` in the obvious way:

```
class Eq a | == a
where
  (<>) infix 4 :: !a !a -> Bool | Eq a
  (<>) x y ::= not (x == y)
```

Here, `<>` is a derived member of the `Eq` class.

Finally, the `Ord` class derives a number of useful operators, once `<` is given, viz. `>`, `<=`, `>=`, `min` and `max`. This makes sense only if `<` is transitive: if $a < b$ and $b < c$, then $a < c$.

3 Do I have to worry about `!`, `*`, `.`, and `u`?:

Clean types are *annotated*, i.e. there is additional information attached to the type of functions, operators, and data types. There are two sorts of annotations: *strictness* (`!`) and *uniqueness* (`*`, `.`, `u`:). To understand the *bare* type of functions and operators, you usually do not need to know the meaning of these annotations, and can therefor safely ignore them. There are two notable exceptions: when working with files (section 4.5) and when working with arrays (section 6). The annotations convey information of the behavior of a function that cannot be derived from its bare type only.

3.1 Strictness annotations

The strictness annotation is the `!` symbol, prefixed immediately before the annotated type. Clean is, by default, a *lazy* language: it evaluates a computation only if it is really needed to obtain the result of the program. Usually, a computation *becomes* needed when it is passed as an argument to a function that in one way or another needs to know (part of) the *value* of that computation. For instance, to compute the sum of two computations, surely `+` requires the values of both its arguments. Another example: to take the head element of a list, surely `hd` must inspect the beginning of the list data structure that it is applied to. This is a property of the function, not of its type. Because types are used to communicate properties of functions in *definition modules*, they are

suited candidates to *piggy back* this information, and therefore we annotate the types of functions:

```
class (+) infixl 6 a :: !a !a -> a    // in StdOverloaded
instance + Int                       // in StdInt
hd :: ![a] -> a                       // in StdList
```

By declaring an instance of `+` for type `Int`, you really make the following function available:

```
(+) infixl 6 :: !Int !Int -> Int
```

Both arguments are annotated with `!`, and hence you know that `+` will have to *evaluate* both arguments in order to compute a sum. Similar, `hd` is defined as:

```
hd [a:x]    = a
hd []       = abort "hd of []"
```

The *pattern match* on the list argument forces the evaluation of any argument passed to `hd` to the structure of the first list constructor, *but no further!* Therefore, it is safe to compute:

```
Start = hd [42 : abort "That hurt!"]
```

3.2 Uniqueness annotations

Clean is a *pure* and *lazy* functional language. To be more precise, it is a pure, lazy *graph rewriting* language, i.e.: it allows *sharing* of arbitrary (sub)computations. This is great from a language engineers point of view because it allows easier reasoning as well as a host of optimizing language transformations. However, can a language that is pure (no assignments) and lazy (no control flow) and that shares computations (long living computations) do interesting things such as file I/O? graphical I/O? use memory efficiently? There is a long answer and a short answer, and I'll stick to the short answer here: yes, they can.

A slightly longer answer is: a function can update an argument *in place* (which is an assignment) *if only* it knows for sure that it is the only piece of code in the program *at the point of evaluation* that has access to that argument. Put in other words: if a function has *unique* access to a data structure, then it can reuse the memory of that data structure without harming any of the highly desirable properties of being pure, lazy and shared.

A function can handle its argument(s) uniquely, i.e.: it does not introduce sharing on that argument. This is a property of the function, and is always

inspected by the *uniqueness analysis* of the **Clean** compiler. However, it is not enough that the function treats this argument uniquely to allow updates on that argument, it must also be certain that that argument *is* unique at every time that the function is called. This can of course not be guaranteed by the function, but only by the calling party. To let them know that this function can update its argument, it annotates the type of that argument with the uniqueness attribute, `*`. This states that the function *will always* update its argument. Examples are opening and closing files, writing data to file, opening windows and menus in a GUI program. In case of *polymorphic* arguments, a function can tell its environment that it does *not change* the uniqueness of that argument by putting the `.` annotation in all occurrences of type argument in its type. This means that you can call the function with either a unique argument or a shared argument. An example is the `hd` function from **StdList**:

```
hd :: ![.a] -> .a
```

You can give it a list that contains elements with shared computations (`[a]`) but also a list in which none of the elements have shared computations (`[*a]`). In either case, `hd` returns the first element of that list, without changing its uniqueness property.

In some cases, these uniqueness dependencies require that you can *name* them so that you can refer to them in other uniqueness constraints. In that case, you can use a name, say `u`, and attach it to a type with `u:`. There are many examples in **StdList**. One such function is `tl`:

```
tl :: !u:[.a] -> u:[.a]
```

This type says that you can apply `tl` to both a list with shared computations as well as a list without shared computations, but the uniqueness of the result list is exactly the same (namely `u`) as that of the argument list.

4 Modules for basic types

Now that we have dealt with overloading and (type constructor) classes as well as annotations (`!`, `*`, `.`, `u:`), we can inspect the modules that define operations on the *basic types* of **Clean**. **Clean** offers six basic types: **Bool**, **Int**, **Real**, **Char**, **File**, and **String**. The **String** type is a bit special because it is really a composite type (array of unboxed **Chars**, to be precise) and is used together with the **StdArray** module, as well as lists, so I treat **Strings** separately in Section 7. For each basic type *Type*, a separate module named **StdType** exists. They consist

mostly of *instance declarations* of the overloaded functions that are defined in the `StdOverloaded` module that was discussed in Section 2.

4.1 StdBool

`StdBool` is the smallest of the basic type modules. Besides a few overloaded operations (`==` and coercion), it provides the usual boolean operators `&&` (conditional and), `||` (conditional or), and `not` (negation). Note that the types of `&&` and `||` suggest that these operations are *conditional*:

```
(||) infixr 2 :: !Bool Bool -> Bool
(&&) infixr 3 :: !Bool Bool -> Bool
```

because the second argument has no strictness annotation `!` (see Section 3.1). If these operations weren't conditional, then they would have to evaluate both arguments in all cases, and therefor would be strict in both arguments.

4.2 StdInt

`StdInt` implements many of the arithmetic overloaded operations that you encounter in `StdOverloaded`. Besides these, it also allows you to do bitwise manipulation of integer numbers. These are:

```
(bitor) infixl 6 :: !Int !Int -> Int
(bitand) infixl 6 :: !Int !Int -> Int
(bitxor) infixl 6 :: !Int !Int -> Int
(<<) infix 7 :: !Int !Int -> Int
(>>) infix 7 :: !Int !Int -> Int
bitnot :: !Int -> Int
```

4.3 StdReal

`StdReal` is very similar to `StdInt` except that it does not offer bit manipulation operations, but instead gives you access to the usual trigonometry operations, as well as raising powers, taking logarithms, and computing the square root.

4.4 StdChar

`StdChar` defines instances for basic computations and coercion on `ascii` characters. Clean does not support Unicode. A few additional coercion functions are defined on the *value* of their `Char` argument:

```

digitToInt :: !Char -> Int    // Convert Digit into Int
toUpper    :: !Char -> Char  // Convert Char into an uppercase Char
toLower    :: !Char -> Char  // Convert Char into a lowercase Char

```

(`digitToInt c`) yields the integer value of the digit representation of that value. Of course, $c \in \{'0' \dots '9'\}$. (`toUpper c`) yields the upper case character if $c \in \{'a' \dots 'z'\}$, and c otherwise. (`toLower c`) yields the lower case character if $c \in \{'A' \dots 'Z'\}$, and c otherwise.

`StdChar` also defines a number of predicates on `Char` values, that are useful when working with texts.

```

isUpper    :: !Char -> Bool   // True if arg1 is an uppercase character
isLower    :: !Char -> Bool   // True if arg1 is a lowercase character
isAlpha    :: !Char -> Bool   // True if arg1 is a letter
isAlphanum :: !Char -> Bool   // True if arg1 is an alphanumerical character
isDigit    :: !Char -> Bool   // True if arg1 is a digit
isOctDigit :: !Char -> Bool   // True if arg1 is a digit
isHexDigit :: !Char -> Bool   // True if arg1 is a digit
isSpace    :: !Char -> Bool   // True if arg1 is a space, tab etc
isControl  :: !Char -> Bool   // True if arg1 is a control character
isPrint    :: !Char -> Bool   // True if arg1 is a printable character
isAscii    :: !Char -> Bool   // True if arg1 is a 7 bit ASCII character

```

4.5 StdFile

`StdFile` is the only module in `StdEnv` that deals with the ‘impure’ external world. `Clean` does offer a lot of other libraries for these kinds of operations, but they are not part of `StdEnv`. The external world is known by the type `World`. There is only one world, so most functions expect that single world and update it. Hence, their type signature is something like $\dots *World \dots \rightarrow \dots *World \dots$. The functions in `StdFile` are grouped into the following categories: managing files, working with read-only files, and working with read-write files. These are explained below.

Managing files: In order to work with a file, you need to *open* it, and *close* it when you’re done. These file management functions are collected within one (type constructor) class, called `FileSystem`:

```
class FileSystem env where
```

```

fopen  :: !{#Char} !Int !*env -> (!Bool,!*File,!*env)
fclose :: !*File      !*env -> (!Bool,      !*env)
stdio  ::              !*env -> (          !*File,!*env)
sfdopen :: !{#Char} !Int !*env -> (!Bool,! File,!*env)

```

In `StdFile`, two instances of this class are provided. The only relevant instance is of type `World`, the other (`Files`) is only present for historic reasons (I can tell you about it if you really want to know). So we have:

```
instance FileSystem World
```

which gives us the following functions to play with:

```

fopen  :: !{#Char} !Int !*World -> (!Bool,!*File,!*World)
sfdopen :: !{#Char} !Int !*World -> (!Bool,! File,!*World)
stdio  ::              !*World -> (          !*File,!*World)
fclose :: !*File      !*World -> (!Bool,      !*World)

```

There are three functions to *open* a file: `fopen`, `sfdopen`, and `stdio`. The difference between `fopen` and `sfdopen` is that `fopen` opens a file that can be written to and read from (it is updated, hence it must be unique), and that `sfdopen` opens a read-only file (it can not be updated, hence it need not be unique). The function `stdio` opens a special file, `stdio`, that connects with the console for simple line-based I/O programs (it is updated, hence it must be unique). Only the unique files need to be closed when you're done with them. This is done with `fclose`. Note that if you accidentally forget to close a unique file, you can not be certain that all data is actually written to that file due to internal buffering. It is also impossible to reopen that file within the same program for further processing.

The first argument of (s)`fopen` is the *file name*. If you simply pass the name of a file, without any directory path before it, then this works on the current directory of the application. If the file name also consists of a directory path, then that location is used. The second argument of both functions controls the *file mode*. This is an integer value, and should be one of:

```

FReadText    := 0 // Read from a text file
FWriteText   := 1 // Write to a text file
FAppendText  := 2 // Append to an existing text file
FReadData    := 3 // Read from a data file
FWriteData   := 4 // Write to a data file
FAppendData  := 5 // Append to an existing data file

```

Use one of the first three modes if you wish to work with *ascii* files, and use one of the last three modes if you wish to work with *binary* files. If you want to *read* either file, use `FReadMode`; if you want to clear the file content and *write* new data, use `FWriteMode`; and if you want to *extend* the current content, use `FAppendMode`. The third parameter must be the current world, of which there is only one anyway.

A few examples: to open a text file, called "test.txt", and read its content:

```
# (ok,file,world) = fopen FReadText "test.txt" world
```

To add data to a log file in some standard directory:

```
# (ok,file,world) = fopen FAppendText
                        "C:\\Program Files\\MyProgram\\settings.log"
                        world
```

`(s)fopen` returns a boolean to report success (`True` only if the file could be opened), the actual file, and the updated world in which the file has been opened. The file result can only be used if the file could be created. Any attempt to use it results in a run-time error, so check your boolean! Hence, typical idiom is:

```
# (ok,file,world) = fopen ...
| not ok           = abort "The file could not opened.\n"
... // now you know it is safe to use file
```

or, if you dislike #:

```
case fopen ... of
  (False,_, world) = ... // opening failed, continue with world
  (True,file,world) = ... // opening succeeded, safe to use file
```

The typical structure of a basic file manipulating program is:

```
Start :: !*World -> *World
```

```
Start world
```

```
# (ok,file,world) = fopen mode name world
| not ok           = abort ("Could not open file '" ++ name ++ "'.")
# file             = do_something_interesting_with file
# (ok,world)       = fclose file world
| not ok           = abort ("Could not close file '" ++ name ++ "'.")
| otherwise        = world
```

```
where
```

```
mode               = // file mode of your choice
name               = // file name of your choice
```

Finally, there is one function that allows you to change the *file mode* of a file that has already been opened with `fopen`:

```
freopen :: !*File !Int -> (!Bool,!*File)
```

The integer argument is the new file mode that the file should have. Again, the boolean returns `True` only if this operation was successful, and the new file value can only be used safely if that was the case.

Read-only files: Read-only files are created with `s fopen`, as explained above. Because they are not updated, they can be shared safely, and hence do not require to be unique. Their type is therefore just `File`.

The following read operations are available on read-only files:

```
sfreadc    :: !File          -> (!Bool,!Char,  !File)
sfreadi    :: !File          -> (!Bool,!Int,   !File)
sfreadr    :: !File          -> (!Bool,!Real,  !File)
sfreads    :: !File !Int     -> (          !*{#Char},!File)
sfreadline :: !File          -> (          !*{#Char},!File)
sfend      :: !File          -> Bool
```

To read the next, single, `Char` (`Int`, `Real`) value from the file, use `sfreadc` (`sfreadi`, `sfreadr`). Because reading advances the internal *read pointer* of the file, these functions need to return a new file value. As usual, the boolean result indicates successful reading.

To read a maximum of `n Char` values from `file`, use `sfreads file n`. To read the next line from `file`, use `sfreadline`. A line is a sequence of `Chars`, returned as a `String`, terminated by a newline character. If the end of the file was not terminated with a newline character, then the `String` simply contains the remainder of the file content. You can test whether you've reached the end of `file` with `sfend file`.

The above functions read a file from start to end. In some cases you want to *seek* a file, i.e. start at specific positions within a file. It is sometimes useful to determine the current value of the internal *read pointer* of a file. This is done with `sfposition file` which returns that value.

```
sfposition :: !File -> Int
```

The internal *read pointer* can be set with `sfseek` which expects three arguments: the first is the file itself and the second and third argument determine the new value of the internal *read pointer*. This third argument determines from which relative position the second argument, an *offset*, should be interpreted:

FSeekSet: set read pointer to *offset*;
FSeekCur: advance read pointer *offset* characters;
FSeekEnd: advance read pointer *offset* characters from the end of the file.

```
sfseek  :: !File !Int !Int -> (!Bool,!File)
```

```
FSeekSet := 0 // New position is the seek offset
FSeekCur := 1 // New position is the current position plus the seek offset
FSeekEnd := 2 // New position is the size of the file plus the seek offset
```

Read-write files: Read-write files support the same read operations as read-only files do. The names of these functions are the same as of the read-only operations with the prefix **s** skipped. The types of the file, which can be updated, is ***File**. Hence, the operations that have been discussed above for read-only files appear similarly for read-write files (the only difference is that **fend** and **fposition** have to return a new file value):

```
freadc  :: !*File          -> (!Bool,!Char,  !*File)
freadi  :: !*File          -> (!Bool,!Int,   !*File)
freadr  :: !*File          -> (!Bool,!Real,  !*File)
freads  :: !*File !Int     -> (          !*{#Char},!*File)
freadline :: !*File        -> (          !*{#Char},!*File)
fend     :: !*File          -> (!Bool,      !*File)
```

```
fposition :: !*File          -> (          !Int,      !*File)
fseek     :: !*File !Int !Int -> (!Bool,      !*File)
```

Besides these operations, which are basically the same as for read-only files, read-write files offer one additional reading function:

```
freadsubstring :: !Int !Int !*{#Char} !*File -> (!Int,!*{#Char},!*File)
```

The meaning of `freadsubstring start end string file` is that `(end-start)` characters are read from `file` (if not possible, then less characters are read). Assume `n` characters are read. These `n` characters are *updated* in `string` (which is why it must be unique), starting at `start`, and ending at `start+n-1`. The result is `n`, the updated string, and the read file.

To write data to read-write files, use one of the following functions:

```
fwritec      ::          !Char    !*File -> *File
```

```

fwritei      ::      !Int      !*File -> *File
fwrite      ::      !Real      !*File -> *File
fwrites      ::      !{#Char} !*File -> *File
fwritesubstring :: !Int !Int !{#Char} !*File -> *File

```

To write a Char (Int, Real, String) to file, use `fwritec` (`fwritei`, `fwrite`, `fwrites`). To write the *slice* `s%(start,end)` to file, use `fwritesubstring start end s file`.

An alternative, and more concise, notation for writing data is the following overloaded operator:

```
class (<<<) infixl a :: !*File !a -> *File
```

```

instance <<< Char
instance <<< Int
instance <<< Real
instance <<< {#Char}

```

Its instance implementations correspond with `fwrite(c,i,r,s)` respectively. Because `<<<` is left-associative, you can write multiple data in one go:

```
file <<< "This line has " <<< 4 <<< " words " <<< '.'
```

Robust programs should check whether file manipulations have succeeded. After you have written data to file, you can check whether this was successful with `ferror file`. Hence, typical idiom of its use is:

```

# file      = file <<< "This line has " <<< 4 <<< " words " <<< '.'
# (fail,file) = ferror file
| fail      = abort "Could not write data file."
... // now you know it is safe to use file

```

If you have created a read-write file, and are completely satisfied, then you can turn it into a read-only file with `fshare :: !*File -> File`. Note that you cannot open this file any longer within the program for further writing!

Finally, there is one special read-write file that is particularly useful for writing *trace statements*:

```
stderr :: *File
```

It is special because it creates a read-write file that can be opened arbitrarily many times in a program. This is not possible for any of the other read-write files, which need to be closed before they can be opened. Note that with the Clean distribution, a more convenient *tracing* module is provided, viz. `StdDebug`.

5 Working with lists

Lists are the workhorses of functional programming languages. The language Clean is no exception to this rule. There is a lot of support for list operations, both in the language as well as in the standard environment (in particular the modules `StdList.dcl` and `StdOrdList.dcl`). This section presents this support.

5.1 List notation

In order to work with lists, you first need to *denote* them. The simplest way of denoting lists is to enumerate all elements. Here are a number of examples:

```
Start = ( []                                     1.
          , [1,2,3,4,5,6,7,8,9,10]             2.
          , [1..10]                             3.
          , ['a'..'z']                           4.
          , ['a', 'c' .. 'z']                   5.
          , [10.0, 9.0 .. 0.0]                 6.
          )                                     7.
```

The simplest list of all is the *empty* list (line 1) which contains no elements. The type of an empty lists cleanly expresses that we do not know its element type: `[] :: [a]`. In line 2 a list is created with 10 elements, viz. 1 upto 10 (inclusive). If there is some sort of regularity, then it is much more concise to use the `..` notation (*dot-dot* notation). Examples are in lines 3 – 6. The list on line 3 has the same value as the one on line 2, and both are of type `[Int]`. The list on line 4 has value

```
['a','b','c','d','e','f','g','h','i','j','k','l','m'
,'n','o','p','q','r','s','t','u','v','w','x','y','z']
```

and type `[Char]`. A less cumbersome notation for `Char`-lists is:

```
['abcdefghijklmnopqrstuvwxyz']
```

which has exactly the same value as above. The examples on lines 5 and 6 shows that lists can also be enumerated that have a regular ‘gap’ between successive elements. The size of the gap is determined by computing the difference between the first two elements. Next elements are computed by keeping adding the gap until the border value has been reached. The list on line 5 has value

```
['acegikmoqsuw']
```

and type `[Char]` and the list on line 6 has value

```
[10.0, 9.0, 8.0, 7.0, 6.0, 5.0, 4.0, 3.0, 2.0, 1.0, 0.0]
```

and type `[Real]`.

If you use *dot-dot* notation, you are required to import the module `StdEnum`. This module is automatically imported if you import `StdEnv`. `StdEnum` contains the functions that create the list when *dot-dot* notation is used. The `Clean` compiler transforms these expressions to these functions. This also explains why it is possible to use *dot-dot* notation for types defined by yourself, as soon as you have defined proper instances of the corresponding class `Enum`.

5.2 ZF-expressions

A powerful way to create and manipulate lists are *ZF-expressions*, or *list comprehensions*. The name *ZF-expressions* comes from the *Zermelo Frank* notation for sets in mathematics. Assume that you have a predicate `isPrimeNumber` which determines whether an integer number is *prime*¹; then the set of all prime numbers can be denoted by:

$$\{x | x \in \{1, 2 \dots\} \wedge isPrimeNumber(x)\}$$

If you have a `Clean` predicate `isPrimeNumber :: Int -> Bool` which determines if a number is prime, then the same set can be constructed as a list by means of a ZF-expression:

```
all_primes :: [Int]
all_primes = [ x \\ x <- [1..] | isPrimeNumber x]
```

Here, the expression `x <- [1..]` is a *generator*. In this case it is the *infinite* list of integral numbers, starting with 1 and incremented by 1. The generator determines the subsequent value of `x`. Just as in function definitions, *conditions* are preceded by a `|` symbol. In this case the condition is `isPrimeNumber x`, for each `x` that is produced by the generator. Hence, the value of the function `all_primes` is an *infinite* list of integral numbers, in ascending order, starting from 1, that are also prime. The result of this function is therefor (at least, its beginning):

```
[1,2,3,5,7,11,13,17,19,23,
```

¹ A number n is prime if its only integral divisors are the numbers 1 and n .

In a ZF-expression the result x can be manipulated as well. Suppose you need the set of squares of all prime numbers, or, mathematically:

$$\{x^2 | x \in \{1, 2, \dots\} \wedge isPrimeNumber(x)\}$$

then you can write this down analogously in Clean as follows:

```
all_prime_squares :: [Int]
all_prime_squares = [ x*x \\< x <- [1..] | isPrimeNumber x]
```

With the other function definition, you could also have defined it as follows:

```
all_prime_squares :: [Int]
all_prime_squares = [ x*x \\< x <- all_primes]
```

One remark on the use of infinite datastructures: in a functional language such as Clean it is no problem to define and manipulate them, as long as you don't try to compute them entirely. That will lead to a non-terminating program. The following function uses this property to compute the first hundred prime numbers:

```
first_100_primes :: [Int]
first_100_primes = take 100 all_primes
```

The standard function `take n xs` takes the first n elements of the list xs and has type `take :: Int [a] -> [a]` (see Section 5.4).

An (inefficient) way to create the function `isPrimeNumber` is the following ZF expression:

```
isPrimeNumber :: Int -> Bool
isPrimeNumber n = [1,n] == [divisor \\< divisor <- [1..n] | n rem divisor == 0]
```

This function examines too many cases to determine whether a number n is prime. However, it should be obvious that it correctly determines the desired property.

5.3 More ZF expressions

In the above section we have created lists with ZF expressions that are constructed with *generators* and *conditions*. It is also possible to combine generators in two ways:

1. If `as` and `bs` are lists, then the *Cartesian product* of both lists can be defined as:

```
product as bs = [(a,b) \\ a <- as, b <- bs]
```

Assume that `as = [1,2,3]` and `bs = ['a', 'b']`, then:

```
product as bs = [(1, 'a'), (1, 'b'), (2, 'a'), (2, 'b'), (3, 'a'), (3, 'b')]
```

The length of the resulting list is equal to the product of the lengths of `as` and `bs`.

2. If `as` and `bs` are lists, then the elements of both lists can be combined element by element (*zipped*):

```
zipped as bs = [(a,b) \\ a <- as & b <- bs]
```

Assume that `as = [1,2,3]` and `bs = ['a', 'b']`, then:

```
zipped as bs = [(1, 'a'), (2, 'b')]
```

The length of the resulting list is the length of the shortest list.

All ways to create and manipulate ZF expressions that have been presented above can be combined in arbitrary ways. As a result, ZF expressions are flexible tools of expression to work with lists.

5.4 StdList: operations on lists

In this section the most frequently used functions of module `StdList` are discussed. These functions are grouped according to what you would like to do with lists.

I want to compare lists:

- `as == bs`: yields `True` only if `as` and `bs` have the same number of elements that also have the same value. Hence, `[a1 ... an] == [b1 ... bm]` is `True` only if `n = m` and $\forall 1 \leq i \leq n : a_i == b_i$.
- `isEmpty as`: yields `True` only if `as = []`.

I want to know how many elements a list has:

- `length [a1 ... an] = n`, and `length [] = 0`.

I want to find an element of a list:

- `isMember x [a0 ... an] = True` only if `x = ai` for some $0 \leq i \leq n$.
- `isAnyMember [x0 ... xn] as = True` only if `isMember xi as` for some $0 \leq i \leq n$.

I want to select an element of a list:

- $[a_0 \dots a_n] !! i = a_i$.
- $\text{hd } [a_0 \dots a_n] = a_0$.
- $\text{last } [a_0 \dots a_n] = a_n$.

I want to remove elements from a list:

- $[a_0 \dots a_n] \% (i, j) = [a_i \dots a_j]$ ($0 \leq i \leq j \leq n$).
- $\text{tl } [a_0 \dots a_n] = [a_1 \dots a_n]$ ($n > 0$).
- $\text{init } [a_0 \dots a_n] = [a_0 \dots a_{n-1}]$ ($n > 0$).
- $\text{take } k [a_0 \dots a_n] = [a_0 \dots a_{k-1}]$ ($0 \leq k \leq n + 1$).
- $\text{takeWhile } p [a_0 \dots a_n] = [a_0 \dots a_k]$ where $(p a_i) = \text{True}$ for all a_i in the resulting list, and $(p a_{k+1}) = \text{False}$.
- $\text{drop } k [a_0 \dots a_n] = [a_k \dots a_n]$ ($0 \leq k$).
- $\text{dropWhile } p [a_0 \dots a_n] = [a_k \dots a_n]$ where $(p a_i) = \text{True}$ for all $i < k$, and $(p a_k) = \text{False}$.
- $\text{filter } p a = [x \setminus x \leftarrow a \mid p a]$.
- $\text{removeAt } i [a_0 \dots a_n] = [a_0 \dots a_{i-1}, a_{i+1} \dots a_n]$.
- $\text{removeMember } x \text{ as}$ removes the first occurrence of x from as .
- $\text{removeMembers } [x_0 \dots x_n] \text{ as}$ removes the first occurrence of $x_0, x_1 \dots x_n$ from as .
- $\text{removeDup } as$ removes all duplicate elements from as (using $==$).
- $\text{removeIndex } e [x_0 \dots x_{i-1} e x_{i+1} \dots x_n] = (i, [x_0 \dots x_{i-1}, x_{i+1} \dots x_n])$, if $e \neq x_j$ voor $0 \leq j < i$.

I want to build a bigger list:

- $[a_0 \dots a_m] ++ [b_0 \dots b_n] = [a_0 \dots a_m, b_0 \dots b_n]$.
- $\text{flatten } [l_0 \dots l_n] = l_0 ++ l_1 ++ \dots ++ l_n$.
- $\text{insertAt } i x [a_0 \dots a_n] = [a_0 \dots a_{i-1}, x, a_i \dots a_n]$.
- $\text{insert } p x [a_0 \dots a_n] = [a_0 \dots a_{i-1}, x, a_i \dots a_n]$, such that $(p x a_i) = \text{True}$, and for each $0 \leq j < i : \neg(p x a_j)$.

I want to modify a list, but retain the number of elements:

- $\text{map } f a = [f x \setminus x \leftarrow a]$.
- $\text{reverse } [a_0 \dots a_n] = [a_n \dots a_0]$.
- $\text{updateAt } i x [a_0 \dots a_n] = [a_0 \dots a_{i-1}, x, a_{i+1} \dots a_n]$.

I want to divide lists:

- `splitAt i a = (take i a, drop i a)`.
- `span p a = (takeWhile p a, dropWhile p a)`.
- `unzip [(a0, b0), (a1, b1) ... (an, bn)] = ([a0, a1 ... an], [b0, b1 ... bn])`.

I want to create a list:

- `iterate f x = [x, f x, f (f x), f (f (f x)) ...]`.
- `repeat x = [x, x, x ...]`.
- `repeatn n x = take n (repeat x)`.
- `indexList [a0 ... an] = [0 ... n]`.
- `zip ([a0, a1 ... an], [b0, b1 ... bn]) = [(a0, b0), (a1, b1) ... (an, bn)]`.
- `zip2 [a0, a1 ... an][b0, b1 ... bn] = [(a0, b0), (a1, b1) ... (an, bn)]`.

I have a list and want to reduce it to a single value:

- `foldl f r [a0 ... an] = f (... (f (f r a0) a1) ...) an`.
- `foldr f r [a0 ... an] = f a0 (f a1 (... (f an r) ...))`.

5.5 StdOrdList: operations on sorted lists

Being sorted is an important property for data structures. The module `StdOrdList` contains functions that manipulate sorted lists. The most frequently used functions are:

- `sort as` sorts the elements of `as` using the ordering `<` that is defined on values of the list element type.
- `merge as bs = sort (as ++ bs)` (but in a smarter way...). Important: `merge` assumes that `as` and `bs` are sorted!
- `maxList as` determines the maximum element of `as` using the ordering `<` that is defined on values of the element type of `as`. `maxList []` yields a run-time error.
- `minList as` determines the minimum element of `as` using the ordering `<` that is defined on values of the element type of `as`. `minList []` yields a run-time error.

Of each of the functions described above, `StdOrdList` also provides a `...By` version, which is parameterized with a comparison function that is used instead of the `<` ordering on values of the list element type.

6 Working with arrays

Clean supports *arrays*. Arrays differ from lists in several aspects:

- Arrays always have a *finite* number of elements, whereas lists can be infinitely long.
- Element selection is in constant time ($\mathcal{O}(1)$), whereas list element selection is linear ($\mathcal{O}(n)$).
- Arrays consume *contiguous* blocks of memory, whereas a list is a linked data structure.

In Clean, the type `String` is implemented as an array of characters. Hence, everything that is said here about arrays also applies to `Strings`. For further manipulations of `Strings`, I refer to section 7.

6.1 Array notation

Arrays are created by enumeration of their elements, in a similar way as with lists, except that the *delimiters* are `{` and `}` instead of `[` and `]`. So, `{}` denotes an array with zero elements, and `{1,2,3,4,5}` and array with five elements, viz. 1 up to 5.

Unfortunately, there is no *dot-dot* notation for arrays, so you can't write down `{1..10}` as you can for lists. However, if you happen to have a list, say `list = [1..10]`, then you can create an array with the same elements as follows:

```
{e \\ e <- list}
```

You can also use another array as a generator. Say you have `array = {1,2,3,4,5}`, then you can create another one:

```
{f e \\ e <-: array}
```

Note the different symbol for extracting elements from an array generator (`<-:`). Because you can use both arrays and lists as generators, you can easily transform lists into arrays and vice versa:

```
toList array = [e \\ e <-: array]
toArray list = {e \\ e <- list}
```

Array creation is *overloaded*: the same expression, say `myArray = {1,2,3,4,5}`, can be one of three concrete types (I refer to the language manual for more details):

- `myArray :: {Int}`, which is a *lazy* array. This is basically an array of unevaluated element expressions.
- `myArray :: {!Int}`, which is a *strict* array. In such an array, all expressions are evaluated strictly.
- `myArray :: {#Int}`, which is an *unboxed* array. Here, the element type must be a basic type. In such an array, the element values are evaluated strictly and stored subsequently within the array.

6.2 Array manipulation

Arrays are overloaded, and this is also the case for its manipulation functions. `Clean` provides syntax for these operations, but internally they are translated to the functions that are imported via `StdArray` (in a similar way as those for *dot-dot* expressions with lists). Because unboxed arrays are only defined for basic types, there must be relation between the sort of array (lazy, strict, unboxed) and its element type. For this reason, the class `Array` that defines these operations is a true type constructor class:

```
class Array .a e where
  select      :: !(a .e) !Int   -> .e
  uselect     :: !u:(a e) !Int  -> *(e, !u:(a e))
  size        :: !(a .e)       -> Int
  useize      :: !u:(a .e)      -> *(!Int, !u:(a .e))
  update      :: !*(a .e) !Int .e -> *(a .e)
  createArray :: !Int e         -> *(a e)
  replace     :: !*(a .e) !Int .e -> *(.e, !*(a .e))
```

The array element selection operations `array.[index]` and `array![index]` are translated to `select array index` and `uselect array index` respectively. The zero-based *index* must be less than the size of the array. If you need an array for subsequent destructive updates, then you should use `uselect`, and otherwise you can use `select`.

To determine the size of a non-unique or a unique array, use `size` and `useize` respectively. Both return the size of the array in constant time.

Unique array updates `{array & [index] = expr}` are translated to `update array index expr`. A useful function is `createArray` which takes an integer that represents the number of elements an array should have, and a value for each of the elements. For instance, `createArray 1000 "Hello World!"` creates a unique array (so that it can be updated destructively later on) that consists of 1000

"Hello World!" text elements. The function `replace` is useful when working with arrays of unique elements: it updates the array at the given index position with the new value, and returns the old value at the same position.

7 Working with texts

As was already mentioned in Section 4, the `String` type is a basic type in `Clean`, even though it is a composite type, viz. an unboxed array of characters. There is special notation for convenient creation (`myString = "Hello World!"`), and there is a module, `StdString`, that contains useful operations on strings. Because strings are also arrays, everything you can do with arrays, you can also do with strings (see Section 6).

`StdString` provides string comparison (`==` and `<`) and the usual conversion instances. Strings can be sliced (`%`) and concatenated (`+++`). If you need to work with unique strings, then `+++` can be used to produce a unique string for further destructive updates. You can update the i th element (counted from zero) of a string s with a new character c with the `:=` operator: `s:=(i,c)`. For instance,

```
Start = let hello = "Hello World!" in hello := (size hello-1, '?')
```

outputs

```
Hello World?
```

Strings are arrays, so the following program produces the same result:

```
Start = {"Hello World!" & [11]='?'}
```

If you need to manipulate strings, it is in many cases more convenient to use lists, simply because `StdList` offers a lot of list manipulating functions. `StdList` contains two functions for this conversion: the `fromString` and `toString` instances for character lists.

Module `StdCharList` offers a number of functions that aim at ‘word processing’ functionality. Usually you get a big chunk of text as a single string, containing newline characters which indicate line ends. Given such a string, you can easily convert it to a list of lines (where each line is a list of characters) with `mklines`. Note that `mklines` removes the trailing newline character from each line. You can then process the line elements, and glue the lines back together into a big string with `flatlines`, which inserts the appropriate newline characters back at the end of each line.

If you need to align lines of text to the left, right or center, given a certain column width, then you can use one of the functions `ljustify`, `rjustify`, or

`cjustify` respectively. As a typical example, the following function transforms a string into another string in which all lines are centered for a certain column width:

```
centerAllLines :: Int -> String -> String
centerAllLines width
  = toString o flatlines o (map (cjustify width)) o mklines o fromString
```

Finally, the little utility function `spaces n` produces a list of `n` spaces (it is literally implemented as `spaces n = repeatn n ' '`).

8 Working with tuples

`StdTuple` contains a number of functions for easy access to tuples and triplets. To select the first or second part of a tuple, use `fst` or `snd` respectively. To select the first or second or third part of a triplet, use `fst3` or `snd3` or `thd3` respectively. Tuples and triplets can be compared for equality (`==` instance) and lexical ordering (`<` instance). If you have two functions f and g , then they can be applied to the first and second component of a tuple with `app2 (f,g)`. Use `app3 (f,g,h)` in case of triplets. Finally, there are two conversion functions to switch between a function of type $(a,b) \rightarrow c$ to $a \rightarrow b \rightarrow c$ (`curry`) and vice versa (`uncurry`).

9 StdFunc

`StdFunc` contains a number of standard higher order functions.

The first two are `id` and `const` with trivial implementations: `id x = x` and `const a b = a`.

When using higher order functions, you sometimes notice that the function that you want to pass to such a function has the arguments in the wrong order. Suppose you have a function `f` of type $a \rightarrow b \rightarrow c$ but need a function in which the argument order is flipped, hence of type $b \rightarrow a \rightarrow c$. Then `flip f` is a function of type $b \rightarrow a \rightarrow c$. Again, `flip` is implemented trivially as `flip f b a = f a b`.

The function composition operator, `o`, is the same as in high school: if you want to apply function g *after* function f , then you can do this with `g o f`. Function composition is concisely defined as: $(o) \ g \ f = \lambda x = g \ (f \ x)$ (note that it must have arity two because it is an operator).

A function that you encounter in literature is `twice`, which is defined as `twice f x = f (f x)`.

`StdFunc` contains a number of functions that resemble imperative control structures:

```
while :: !(a -> .Bool) (a -> a)  a -> a
until :: !(a -> .Bool) (a -> a)  a -> a
iter  :: !Int          (.a -> .a) .a -> .a
```

Given a predicate `p`, and a computation `f` on some arbitrary data type, and an initial value `v0`, `while p f v0` applies `f` to subsequent values $v_i = f^i v_0$ while `p` holds for these values. The value returned is the first v_n for which `p vn` is false. The computation `until p f v0` is basically the same, except that the computations are performed as long as `p` does not hold, and the value returned is the first v_n for which `p vn` is true. The computation `iter n f v0 = fn v0`, provided that $n \geq 0$.

The function `seq` is given a list of computations, $[f_0, \dots, f_n]$ that all modify a value of some arbitrary type, and computes their subsequent application in left-to-right order: $f_n \circ f_{n-1} \circ \dots \circ f_0$. The function `seqList` is similar, except that the functions in the list not only modify a value of some arbitrary type, but also return some additional information. `seqList` applies the functions in the list in the same order as `seq` does, but also collects these additional results in the same order, and returns them together with the altered value. Both functions are most often used in combination with unique data structures, such as files.

The final two functions are extremely well known in literature, and they are known as the *monadic state combinators* (for clarity, I show the types here without uniqueness annotations):

```
(‘bind’) infix 0 :: (St s a) (a -> St s b) -> St s b
(‘bind’) f g = \st0 -> let (result,st1) = f st0 in g result st1
```

```
return :: a -> St s a
return r = \s -> (r,s)
```

`bind` is a kind of ‘flipped’ function composition, except that the subsequent function can ‘react’ to the result of the first computation. `return` leaves the state untouched, and returns it as well as its argument value. These two higher order functions are extremely useful as glue of computations that return additional information, as we have also seen in `seqList`. You encounter it typically in combination with λ abstraction, in which you use the λ to give a name to the result of the first computation, and continue with the second computation that can use that name during its computation.

10 StdMisc

The smallest module of StdEnv is `StdMisc`, which defines two functions that are useful to indicate error-conditions in a program:

```
abort :: !{#Char} -> .a
undef ::                .a
```

As you can see from their types, they can be applied in any context. `abort msg`, when evaluated, prints `msg` and aborts the computation of the entire program, while `undef` simply aborts without an additional message. Use these functions to cover unexpected cases in your code.