



The Generic HASKELL User's Guide

Version 1.42 (Coral)

The Generic HASKELL Team

Andres Löh (editor)
Johan Jeuring (editor)

Dave Clarke
Ralf Hinze
Alexey Rodriguez
Jan de Wit

info@generic-haskell.org

14 January 2005

Institute of Information and Computing Sciences
Utrecht University
P.O. Box 80.089
3508 TB Utrecht
The Netherlands
<http://www.generic-haskell.org>

Contents

1	What is Generic HVSHELL?	5
1.1	Generic programming	5
1.2	Generic HVSHELL overview	5
2	Installation	6
2.1	System requirements	6
2.2	Installing the binary distribution (Linux, Mac OS X)	6
2.3	Installing the binary distribution (Windows)	7
2.4	Building from source	7
2.5	Running gh	7
2.6	Command line flags	8
2.7	General overview of compilation	9
2.8	Compiling and running the generated code	9
3	Generic HVSHELL: The Language	10
3.1	Special Parentheses	10
3.2	Type-indexed values	10
3.3	Generic type signatures	13
3.4	Generic application	15
3.5	Local redefinition	16
3.6	Default cases	16
3.7	Generic abstraction	17
3.8	Type-indexed types	18
3.9	Specialisation	19
3.10	Generated function naming	19
3.11	Module system	19
3.12	Haskell compatibility	20
4	Library	21
4.1	Introduction	21
4.2	Module <code>GH.Library.Bounds</code>	21
4.3	Module <code>GH.Library.Collect</code>	21
4.4	Module <code>GH.Library.Compare</code>	22
4.5	Module <code>GH.Library.DeepSeq</code>	22
4.6	Module <code>GH.Library.Enum</code>	23
4.7	Module <code>GH.Library.Eq</code>	23
4.8	Module <code>GH.Library.Map</code>	23

4.9	Module <code>GH.Library.MapM</code>	24
4.10	Module <code>GH.Library.ReadShow</code>	24
4.11	Module <code>GH.Library.Reduce</code>	25
4.12	Module <code>GH.Library.Table</code>	25
4.13	Module <code>GH.Library.ZipWith</code>	26
5	Future Work	27
6	Meta-information	28
6.1	Contact	28
6.2	Caveats	28
6.3	Known bugs and limitations	28
6.4	Change log	29
6.5	Acknowledgements	29
6.6	Copyright information	29

1 What is Generic HASKELL?

1.1 Generic programming

Software development often consists of designing datatypes around which functionality is added. Some functionality is datatype specific, whereas other functionality is defined on almost all datatypes in a way that depends only on the structure of the datatype. A function that works on many datatypes in this manner is called a *generic* (or polytypic) *function*. Examples of generic functionality include editing, pretty-printing or storing a value in a database, and comparing two values for equality.

Since datatypes often change and new datatypes are introduced, we have developed Generic HASKELL, an extension of the functional programming language Haskell [10] that supports generic definitions, to save the programmer from (re)writing instances of generic functions. The original design of Generic HASKELL is based on work by Ralf Hinze [3]. The current release is based on recent work by Dave Clarke, Johan Jeuring and Andres Löh [9, 8]. It extends Haskell with, among other things, a construct for defining type-indexed values. These values can be specialised to all Haskell datatypes, facilitating wider application of generic programming than provided by earlier systems such as PolyP [7].

1.2 Generic HASKELL overview

Generic HASKELL extends Haskell with a number of features.

- *type-indexed values* are defined as values indexed over the various Haskell type constructors (unit, primitive types, sums, products, and user-defined type constructors). In addition, we can also specify the behaviour of a type-indexed value for a specific constructor using *constructor cases*, and reuse one generic definition in another using *default cases*.

The resulting type-indexed value can be specialised to any type.

- *type-indexed types* are types which are indexed over the type constructors. These can be used to give types to more involved generic values. The resulting type-indexed types can be specialised to any type.
- generic definitions can be used by applying them to a type. This is called *generic application*. The result is a type or value, depending on which sort of generic definition is applied.
- *generic abstraction* enables generic definitions to be defined by abstracting a type parameter (of a given kind).

2 Installation

2.1 System requirements

Generic HVSHELL is written in Haskell. The package has been tested with GHC and comes with a `Makefile` suitable to build it using GHC. Parts of the system are written using Utrecht University's attribute grammar system `ag`, Ralf Hinze's `frown :-()` parser generator, and the DrIFT tool, but neither of these tools is required to build the compiler, because the generated Haskell sources are included in the distribution. The code that is generated by Generic HVSHELL is such that it compiles using GHC, with extensions enabled. The generated code makes use of rank- n polymorphism and of explicit kind-annotations for data types – otherwise, it is Haskell 98 compliant.

Two kinds of distribution are available.

The *binary* distribution includes a `gh` compiler binary. The compiler translates Generic HVSHELL input files into Haskell files, thus GHC is still required. We currently provide binaries for Linux, Windows, and Mac OS X. In principle, we can provide binaries for any platform supported by GHC.

The *source* distribution includes the Haskell code for the `gh` compiler which has been generated from our compiler source using both `ag`, `frown :-()`, and DrIFT. GHC is required to compile this distribution. Configuration files are provided.

Current development versions and snapshots can be obtained by accessing the Generic HVSHELL Subversion repository at

<https://svn.cs.uu.nl:12443/repoman/info/Generic-Haskell>.

2.2 Installing the binary distribution (Linux, Mac OS X)

Installation is also explained in the `INSTALL` file included in the distribution.

1. Unpack the tarball to a directory of your choice and change to this directory.
2. Issue the following commands:

```
cd build
../configure --prefix=install_path
```

The flag `--prefix=install_path` is optional and defaults to `/usr/local/`. It can be used to select the place in the filesystem where the files specific to Generic HVSHELL will be installed.

3. Next run (GNU make is required):

```
make install
make package
```

The first command will move all the files to their final locations. In particular, the binary `gh` will be placed in the directory `${prefix}/bin`. The second command will register the package `generic-haskell` – containing the Generic HVSHELL libraries – with GHC.

Note that `make install` requires that you have write permissions to the installation directories, and `make package` requires that you have write permission to the global GHC package configuration file.

Alternatively, you can install Generic HVSHELL for usage “in-place”, using the command

```
make in-place
```

4. If you have chosen a global installation, you can remove the distribution directory now. For an in-place installation, it is still required.

2.3 Installing the binary distribution (Windows)

Installation is also explained in the `INSTALL` file included in the distribution.

1. Executing the distribution file will start an installation assistant which will guide you through Generic HVSHELL installation.
2. Follow the instructions of the assistant.

2.4 Building from source

Building from source requires exactly the same command sequence as for installing the binary distribution on a Unix system. The difference of course is the amount of work which is subsequently done.

If you are on a Windows system, Cygwin or a similar system is required to be able to use the Makefile.

2.5 Running gh

The Generic HVSHELL compiler is called `gh`. It has essentially two modes of operation.

- If you call `gh` without supplying a file to process, you will be asked for a file name. Specify an input file relative to your working directory, and the compiler will process the file and generate an output file with the same basename as the input file, but the extension `.hs`. It will also generate an interface file with the extension `.ghi`.
- Alternatively, input files can be specified on the command line.

A typical invocation is:

```
gh your_file.ghs
```

Again, the compiler will produce an `.hs` and a `.ghi` file as result.

2.6 Command line flags

A number of command line flags are available:

```
Usage: gh [options...] files...
  -v      --verbose           (number of v's controls the verbosity)
  -m      --make              compile dependent files first
  -V      --version, --release show version info
  -h, -?  --help              show help
  -C      --continue          continue after errors
  -L DIR  --library=DIR       add DIR to search path
                          --ghc           use GHC extensions in output (default)
                          --no-ghc        use no GHC extensions in output
                          --odir=DIR      place output files in DIR
                          --print-searchpath print search path
```

Verbosity (-v) The amount of output generated by the compiler can be selected by providing a number of `-v` flags. The default level of verbosity (no `-v` flag) produces only error messages. The second level (`-v`) in addition provides some diagnostic information and warnings. The third and fourth level (`-vv` and `-vvv`) in addition provide a significant amount of debugging information.

Track module dependencies (-m) The `-m` (or `--make`) option attempts to chase dependencies and compile those which require compilation. There are unfortunately cases where this fails.

Continue after errors (-C) The `-C` (or `--continue`) option forces the compiler to continue compilation even when an error is encountered. This can be used to generate more than one error message or to see the resulting generated code, but unfortunately may result in the compiler running until it crashes.

Search path (-L) The Generic HASKELL compiler needs to find a number of files, in particular the Generic HASKELL prelude. There are three possibilities to make the location of the standard libraries known to the compiler:

- Set the environment variable `GH_HOME` to the directory you unpacked the Generic HASKELL distribution to.
- Set the environment variable `GH_LIBRARY_PATH` to the directory where the libraries are located (usually `GH_HOME/lib`).
- Pass the path where the libraries are located as an argument to the compiler, using the `-L` option. This option can also be used to add other directories to the search path.

During installation, `GH_HOME` is preset to the installation directory, and the `gh` executable is in fact a wrapper script first setting the `GH_HOME` variable and then calling the real binary called `gh-bin`. If `GH_HOME` is set by the user, this information overrides the preset directory.

For debugging purposes, the search path used by Generic HVSKELL can be queried using the `--print-searchpath` flag.

Language features used (`--ghc`, `--no-ghc`) By default, generated code makes use of explicit kind annotations in datatypes declarations, as supported by GHC when extensions are enabled via `-fglasgow-exts`. If you want to try running Hugs on the generated files, you should pass the flag `--no-ghc` to the Generic HVSKELL compiler, which will prevent explicit kind annotations from being generated, at the risk that the Haskell compiler will subsequently complain about kind errors.

2.7 General overview of compilation

The Generic HVSKELL compiler compiles `.ghs` files and produces `.hs` files which can subsequently be compiled using a Haskell compiler. In addition, the compiler also produces `.ghi` interface files for compiled modules, which will be used in subsequent compilations to avoid unnecessary recompilation.

2.8 Compiling and running the generated code

The Generic HVSKELL compiler generates ordinary Haskell code which can be run or compiled using GHC. The Generic HVSKELL distribution includes wrapper scripts called `gh-ghc`, `gh-ghci` and `gh-hugs`, for calling GHC, GHCi and Hugs, respectively. These wrappers automatically pass some options appropriate for running programs generated by Generic HVSKELL, such as making the libraries accessible.

Note that Hugs is not generally supported as a Generic HVSKELL backend. If you want to try running your Generic HVSKELL programs in Hugs, you should use the `--no-ghc` flag when calling `gh`.

3 Generic HASKELL: The Language

The Generic HASKELL compiler implements a number of extensions to Haskell. These are described briefly here, but this chapter is not intended as an introduction to generic programming. Further information can be found by consulting the literature [8, 4, 5, 2, for example]. In particular, Exploring Generic Haskell (EGH) [8] explains the Generic HASKELL language in much more detail than can be covered here. Most of the syntax used in EGH can be used in the Generic HASKELL compiler. Where there are significant differences, we point them out in this guide.

Furthermore, the program files included in the distribution (the `.ghs` files in the subdirectories `lib/`, `examples/`, and `tests/`) contain several examples of generic programming.

3.1 Special Parentheses

Type-indexed definitions take a type argument which is surrounded by special parentheses. The parentheses $\{\{_ _ \}\}$ (i.e., $\{\{ _ _ \}$) wrap such a type argument. Although no longer required, we still support the parentheses $\{\{_ _ \}$ (i.e., $\{\{ _ _ \}$), wrapping a kind argument of a kind-indexed type.

3.2 Type-indexed values

Generic HASKELL introduces a new top-level declaration for type-indexed values. A type-indexed value is defined using the following syntax:

```
 $\langle \text{varid} \rangle \{\{\langle \text{gtyvars} \rangle\}\} \quad :: \langle \text{gtypesig} \rangle$   
 $\langle \text{varid} \rangle \{\{\langle \text{typat} \rangle\}\} \quad \dots = \dots$   
 $\langle \text{varid} \rangle \{\{\langle \text{typat} \rangle\}\} \quad \dots = \dots$   
:
```

A type-indexed value consists of a type signature followed by one or more cases that define the function by matching on a type argument. Let us first look at these cases.

Type patterns A type pattern $\{\{\langle \text{typat} \rangle\}\}$ usually is the name of a type constructor, (fully) applied to distinct type variables, so that the resulting pattern is a type expression of kind `*`. For example,

```
Unit  
[a]  
Either a b  
a → b
```

are all valid type patterns, whereas

```
Either a -- missing the second type variable
a → a -- no distinct variables
Int → a -- (→) is not applied to only variables
```

are all illegal.

Type patterns may also consist only of the name of a type constructor. Thus,

```
Either
Sum
(→)
```

are also valid type patterns. Such *named types* correspond to the style of generic programming that was supported by **Generic HASKELL** prior to the Coral release. If named types of a kind other than `*` are used as type patterns, then the case of the function may take additional parameters according to the dependencies of the function. **Generic HASKELL** now also supports *Dependency-style* definitions, which make use of the type patterns with type variables.

Structural representation types Type-indexed values become generic if they are defined for some or all of the structural representation types, which are used internally by **Generic HASKELL** to represent Haskell datatypes. These types are

```
data Unit = Unit
data Sum a b = Inl a | Inr b
-- Sum can be written as :+: in type indices
data Prod a b = a *: b
-- Prod can be written as *: in type indices
data Con a = Con a
data Label a = Label a
```

These types are defined in the module `GH.Prelude`, which is automatically imported in each **Generic HASKELL** program. The types `Con` and `Label` are used to represent constructors and record field labels, respectively. When used as a type pattern, they get an additional argument which is bound to a *value* (not a type) of type `ConDescr` or `LabelDescr`.

The correct *Dependency-style* type patterns for `Con` and `Label` are thus:

```
Con d a
Label d a
```

and as named types, they receive the additional argument as well:

```
Con d
Label d
```

The datatypes `ConDescr` and `LabelDescr` are given here. These can be used for querying information about constructors and labels.

```

data ConDescr = ConDescr { conName   :: String,
                           conType   :: String,
                           conArity  :: Int,
                           conPosition :: (Int, Int),
                           conLabels  :: Bool,
                           conFixity  :: Fixity }

data Fixity      = Nonfix
                  | Infix  { prec :: Int }
                  | Infixl { prec :: Int }
                  | Infixr { prec :: Int }

data LabelDescr = LabelDescr { labelName :: String,
                                labelType  :: String,
                                labelStrict :: Bool }

```

Naturally, none of these predefined identifiers should be used in the remainder of a program in a way that clashes with their use in generic definitions, following the usual scoping rules of Haskell.

Example A type-indexed value that checks whether two given values are structurally equal can be defined as follows:

```

eq {a} :: a → a → Bool
eq {Unit}   Unit   Unit   = True
eq {Sum a b} (Inl x)  (Inl y)  = eq {a} x y
eq {Sum a b} (Inr x)  (Inr y)  = eq {b} x y
eq {Sum a b} _         _         = False
eq {Prod a b} (x1 :* x2) (y1 :* y2) = eq {a} x1 y1 ∧ eq {b} x2 y2
eq {Int}      m         n         = m == n   -- Haskell standard equality
eq {Char}     c         d         = c == d   -- Haskell standard equality

```

Generics defined for user-defined types Type-indexed values (and also type-indexed types) can be defined over (possibly user-defined) datatypes, not only the structural representation types. This covers the cases for Int and Char, shown above. Additional cases such as the following are also possible:

```

eq {Range}   (R _ _) (R _ _) = True

```

for the user defined type

```

data Range = R Int Int

```

Constructor cases Datatypes which have a large number of constructors often require functions that behave in some uniform manner for most constructors, but in some specific way for certain other constructors. To write such functions **Generic HASKELL** allows cases for specific constructors to be written. Using these *constructor cases* a generic function can have special cases to deal with the constructors requiring special treatment.

The syntax of the case is given by

```
case <qcon>
```

as illustrated in the following

```
freecollect {case Lambda} (Lambda (v, t) e) = filter (≠ v) (freecollect {Expr} e)
```

The case `freecollect {case Lambda}` will be applied only when the value of type `Expr` (from which `Lambda` is a constructor) encountered has the form `Lambda (v, t) e`. The case should be written to exploit this knowledge. Interestingly, when a constructor case produces a value, it need not produce a value with the same constructor, but only of the correct type.

The type of a constructor case is the same as the type from which the constructor comes. Thus, since `Lambda` is a constructor for the `Expr` datatype, the type of the right-hand side is what it would be for the `Expr` case.

3.3 Generic type signatures

The syntax of a generic type signature is as follows:

```
<varid> {<gtyvars>} :: <gtypesig>
```

Generic HVSHELL supports two styles of generic type signatures. New in Coral, it supports Dependency-style Generic HVSHELL. A Dependency-style type signature uses one or more *generic* and zero or more *non-generic* (or *parametric*) type variables, which are separated using a vertical bar, as type pattern. The pair of integers consisting of the number of generic and parametric type variables of a type-indexed value is called its *arity*. Generic type variables must always be of kind `*`, whereas parametric type variables may have different kinds.

The example signature of the equality function,

```
eq {a} :: a → a → Bool
```

consists of one generic variable `a`, and no parametric variables. Other examples are the signatures of `map` and `collect`,

```
map {a1, a2} :: (map {a1, a2}) ⇒ a1 → a2
collect {a | v} :: (collect {a | v}) ⇒ a → [v]
```

with two generic and one generic plus one parametric variable, respectively. The precise meaning of generic and parametric type variables is discussed in EGH. Other than in EGH, the kind of the type variables can be omitted if it can be inferred.

The right-hand side of a Dependency-style type signature is of the form

```
(<deps>) ⇒ (<type>)
```

a list of dependencies, followed by the *base type*.

Dependencies If a type-indexed function refers directly or indirectly to another type-indexed function in its definition, it is usually (exceptions: the type argument of the call is constant, or the other function is a generic abstraction) a *dependency* of the function.

The equality function, as given in the example, depends on itself, so the full type signature of *eq* is

```
eq {a} :: (eq {a}) => a -> a -> Bool
```

Often, **Generic HVSHELL** can infer dependencies. If no list of dependencies is given in a generic type signature, **Generic HVSHELL** will try to infer the dependencies and complain if it does not succeed. Dependency inference is a rather experimental feature, and it is likely to be changed in future releases.

Note that in **Generic HVSHELL** a dependency in a type signature must always include the variables, the EGH abbreviation of omitting the variables for functions of arity $\{1 \mid 0\}$ is not available.

In older releases, dependencies of generic functions had to be specified explicitly using the **dependency** keyword. This is still possible when generic functions are defined using kind-indexed types, but not needed when Dependency-style type signatures are used.

Kind-indexed types In previous releases of **Generic HVSHELL**, type signatures of generic functions made use of kind-indexed types. This style of generic programming is still supported. In fact, Dependency-style type signatures are internally converted into applications of kind-indexed types. Kind-indexed types are thus no longer required, but still supported for backwards compatibility.

The function *eq* with the type signature given above behaves as if it had been defined using the kind-indexed type *Eq*:

```
type Eq {[*]}      a = a -> a -> Bool
type Eq {k -> l} a =
  forall u . Eq {k} u -> Eq {l} (a u)
eq {t :: k} :: Eq {k} t
```

Similarly, the function *collect* could be defined using the following kind-indexed type *Collect*:

```
type Collect {[*]} a v = a -> [v]
type Collect {k -> l} a v =
  forall u . Collect {k} u v -> Collect {l} (a u) v
collect {t :: k} :: forall v . Collect {k} t v
```

Note that the type variable *v* that is parametric in the Dependency-style type signature of *collect* is just passed on unchanged in the kind-indexed type, and is universally quantified at the outer level.

If a function depends on other functions except itself, a kind-indexed type – in contrast to a Dependency-style type signature – is not sufficient. The dependencies have to be made explicit using the **dependency** keyword. For example, the hypothetical function

```
pretty {a} :: (important {a}, pretty {a}) => a -> String
```

which pretty-prints the important parts of a data structure, can also be defined using the kind-indexed type

```

type Pretty  $\{\{*\}\}$   $a = a \rightarrow \text{String}$ 
type Pretty  $\{\{k \rightarrow l\}\}$   $a =$ 
  forall  $u$ . Important  $\{\{k\}\}$   $u \rightarrow$  Pretty  $\{\{k\}\}$   $u \rightarrow$  Pretty  $\{\{l\}\}$   $(a\ u)$ 
pretty  $\{\{t :: k\}\}$  :: Pretty  $\{\{k\}\}$   $t$ 

```

but then, the statement

```

dependency pretty  $\leftarrow$  important pretty

```

must be supplied in addition. In general, kind-indexed types are defined according to the syntax:

```

type  $\langle \text{Conid} \rangle$   $\{\{*\}\}$   $t_1 \dots t_n = \langle \text{type} \rangle$ 
type  $\langle \text{Conid} \rangle$   $\{\{k \rightarrow l\}\}$   $t_1 \dots t_n = \langle \text{type} \rangle$ 

```

3.4 Generic application

A type-indexed value can be specialised to a value by applying it to a type. Generic application extends the syntax of expressions ($\langle aexp \rangle$) as follows:

```

 $\langle aexp \rangle ::= \dots$ 
  |  $\langle \text{varid} \rangle \{\{ \langle \text{type} \rangle \}\}$ 

```

The type argument must not contain universal quantifiers or class constraints. When a generic function is used, its type argument must always be supplied – it cannot (yet) be inferred.

Similarly, a kind-indexed type can be specialised to a type by supplying the kind at which the definition is to be applied. The syntax of type expressions ($\langle gtycon \rangle$) is thus extended as follows:

```

 $\langle gtycon \rangle ::= \dots$ 
  |  $\langle \text{Conid} \rangle \{\{ \langle \text{kind} \rangle \}\}$ 

```

Example Given the datatype:

```

data BinTree  $a = \text{Empty} \mid \text{Node } a \text{ (BinTree } a \text{) (BinTree } a \text{)}$ 

```

The *map* function for BinTree is $\text{map } \{\{\text{BinTree}\}\}$. The type of $\text{map } \{\{\text{BinTree}\}\}$ is $(a \rightarrow b) \rightarrow (\text{BinTree } a \rightarrow \text{BinTree } b)$.

Short notation Type-indexed values (that are not defined via generic abstraction, see Section 3.7) can be applied to type arguments of any kind. The dependencies for the type arguments are required in the same order as specified in the type signature or the **dependency** statement for the function. This mechanism is called *short notation* in EGH.

Other than in EGH, it is currently possible to use short notation with functions that have multiple dependencies. It is even possible to use short notation with functions that have inferred dependencies, although this is not recommended, as the order of dependencies then depends on the implementation of the inference algorithm, and this behaviour is likely to be disabled in future releases.

3.5 Local redefinition

It is possible to locally redefine the behaviour of a generic function, such as described in Chapter 8 of EGH. Local redefinition looks much like an ordinary **let** statement:

```
let <varid> {a0 a1 ... an} ... = ...
  ⋮
in ...
```

The generic function referred to by *<varid>* must be in scope. Here, a_0, a_1, \dots, a_n are all type variables. The variable a_0 is the variable on which the function is redefined, the arguments a_1, \dots, a_n are local to the right-hand side, and only required if a_0 is of kind other than $*$.

Example Assume that *size* is a generic function of type

```
size {t} :: (size {t}) ⇒ t → Int
```

computing the size of a value. All base types and Unit return 0. In sums, the size of the particular alternative is chosen, and in products, the sizes of the components are added. This function is useless on types of kind $*$ without local redefinition, because it would always return 0. However,

```
let size {a} = const 1
in (size {[Int]} [[1, 2, 3], [4, 5]],
    size {[a]} [[1, 2, 3], [4, 5]],
    size {[a]} [[1, 2, 3], [4, 5]],
    size {a} [[1, 2, 3], [4, 5]])
```

evaluates to (0, 5, 2, 1).

3.6 Default cases

Default cases allow one generic definition to be defined by implicitly copying the lines from another, updating and adding cases where appropriate. This is particularly useful for defining functions which follow a specific traversal pattern.

Note Between the Beryl and Coral releases, the syntax of default cases has been changed in a non backwards-compatible way. The original syntax [1] clashes with the more liberal syntax of generic abstractions and has therefore been removed. Instead, the syntax described in Chapter 14 of EGH is supported.

Example Suppose we have a crush-like function which collects a list of values of type a from some datatype.

```
collect {a | v} :: (collect {a | v}) ⇒ a → [v]
```

We can adapt this function to collect values of type `Var`, to produced a function of the following more specific type

```
varcollect {a} :: (varcollect {a | Var}) ⇒ a → [Var]
```

by writing but a few lines:

```
varcollect extends collect
varcollect {Var} v = [v]
varcollect {Prod a b} (x:*:y) = varcollect {a} x ∪ varcollect {b} y
```

The line containing the keyword **extends** is the default case, which has the effect of copying the code from `collect` into the new generic function `varcollect`. The line for `varcollect {Var}` specifies the desired additional functionality for type `Var`. The line for `varcollect {Prod a b}` overrides the functionality for `Prod`, using union instead of concatenation to accumulate the results.

3.7 Generic abstraction

A type variable (of fixed kind) can be abstracted generically from an expression using a *generic abstraction*. Declarations take the following form:

```
⟨varid⟩ {⟨gtyvars⟩} :: ⟨gtypesig⟩
⟨varid⟩ {t} ... = ⟨exp⟩
```

The difference with ordinary type-indexed values is that the generic type arguments in the type signature may be of a kind other than `*`, and the type variable t must be of the same kind. Furthermore, a generic abstraction has exactly one case, of the given form.

Example An example is the so-called categorical strength:

```
strength {t} :: t a → b → t (a, b)
strength {t} ta b = map {t} (λx → (x, b)) ta
```

Deviating from EGH, the dependencies of generic abstractions are usually inferred when no explicit dependencies are provided in a type signature. The same holds for the kind of the type argument. It is, however, possible, to specify everything explicitly:

```
strength {t :: * → *} :: (map {t, t}) ⇒ t a → b → t (a, b)
```

3.8 Type-indexed types

Warning The implementation of type-indexed types has not been updated to support Dependency-style and the syntax used in EGH. This will probably be implemented in the next release of Generic HVSHELL. At the moment, the implementation of type-indexed types is essentially the same as in the Beryl release.

Type-indexed types [6] can be defined just as type-indexed values, except that the right-hand side of a definition is a constructor followed by a type. Thus the syntax consists of a collection of definitions, indexed over the type constructors, of the form:

```
type <Conid> {<stype>} t1 ... tn = <con> <type>
```

Only named type constructors are supported as <stype>, and Con and Label do not get additional descriptor arguments on the type level. New constructors (<con>) must be introduced for each case of such a definition – each case will be compiled into a **newtype** declaration.

A type-indexed type can be specialised to a type by supplying its type argument.

```
<gtycon> ::= ...
           | <Conid> {<type>}
```

Example The type-indexed type FMap is defined as follows:

```
type FMap {Unit}      v = FMU (Maybe v)
type FMap {Sum} fma fmb v = FMP (fma v, fmb v)
type FMap {Prod} fma fmb v = FMT (fma (fmb v))
type FMap {Con} fm      v = FMC (fm v)
type FMap {Label}     v = FML (fm v)
```

FMap can be used anywhere a type can by supplying FMap with a type parameter, for example in the following:

```
type Lookup {[*]} t = forall v . FMap {t} v → t → Maybe v
type Lookup {k → l} t = forall a . Lookup {k} a → Lookup {l} (t a)
```

The constructors introduced in the definition of FMap can be used in pattern matching:

```
lookup {t :: k}                :: Lookup {k} t
lookup {Unit} (FMU fm) Unit    = fm
lookup {Sum} lA lB (FMP (fma, fmb)) (Inl a) = lA fma a
lookup {Sum} lA lB (FMP (fma, fmb)) (Inr b) = lB fmb b
lookup {Prod} lA lB (FMT fma) (a :: b) = do fmb ← lA fma a
                                         lB fmb b
lookup {Con d} l (FMC fm) (Con b) = l fm b
lookup {Label d} l (FML fm) (Label b) = l fm b
```

3.9 Specialisation

Generic functions are specialised at compile time, thus no run-time representation of types is required. (There is however the cost of encoding and decoding types.) The compiler determines which specific types a generic function is used with, and then generates the set of specialised versions for that function in the output file.

Specialisations are always generated locally per module. Thus, a generic function which is defined in one module but used in many, results in some work being duplicated.

The compiler proceeds by collecting *specialisation requests* and *implications* from the source. The implications are then applied to the requests repeatedly, yielding new requests, until a fixpoint is reached.

If the compiler is working correctly, the fixpoint calculation always terminates.

3.10 Generated function naming

The Generic HVSHELL programmer must be aware that the generated Haskell code is polluted with additional names corresponding to instances of generic functions. These may clash with a programmer's own function names. Fortunately, this is highly unlikely as the generated names are rather complicated, encoding details such as module and type names. Names generated by the compiler all begin with `gh?_` or `GH?_`, where `?` is an arbitrary letter.

Unfortunately, this obfuscation makes it difficult to directly interface ordinary Haskell code with the code generated by the Generic HVSHELL compiler. We offer a tip to the adventurous who wish to do such a thing. If you wish to use a generic function such as `map {List}` in ordinary Haskell code, add a line such as

```
mapList = map {List}
```

to the appropriate Generic HVSHELL file, and then use the function `mapList` in your Haskell code.

3.11 Module system

The module system of Generic HVSHELL mirrors the behaviour of Haskell's module system, as far as the Haskell language is concerned. Additionally, generic entities (i.e., kind-indexed types, type-indexed values, and type-indexed types) may appear in export and import lists. If no export or import list is given, then all generic entities are exported or imported, respectively. If a generic entity appears in a list, then all of its cases are exported or imported. It is not possible to export only some cases of a type-indexed value, or to limit the constructors visible for a type-indexed type.

It is recommended that the kind-indexed type of a type-indexed value is also exported. Forgetting to do so may result in unexpected behaviour.

At the moment, it is not possible to define a type-indexed value or type-indexed type across modules. However, one can achieve a similar effect by importing a generic function qualified and redefining a new function with the same name by means of a default case.

3.12 Haskell compatibility

Generic HASKELL parses all Haskell programs, except in the following instances:

- The tokens **forall**, **extends**, and **dependency** are additional keywords in Generic HASKELL.
- The special parentheses for type and kind arguments, i.e., `{|}`, `{[}`, are all handled as a single token. Unfortunately, some pieces of regular Haskell code can trick the lexer and result in parse errors. For example, in

```
do { [x] ← action; return x }
```

the initial `{[` is treated as a single token `{[` rather than the two tokens `{` and `[` which an Haskell programmer would expect. In other instances, sequences such as `+|}` are considered as the operator `+|` followed by a `}`, since `|` may occur in operators, whereas `{|+` is considered as the token `{|` followed by `+`.

The required fix in both cases is to insert a space in the appropriate place, for example, by writing instead

```
do { □ [x] ← action; return x }
```

4 Library

4.1 Introduction

The Generic HVSHELL system comes with a library of useful generic functions. These are summarised below; for the details, consult the library itself (in subdirectory `lib`). We give the Dependency-style type signatures of the generic functions, as well as the instantiated types for kind `*` and kind `* → *`, and usually a short description.

Naming conventions When generic functions defined in the Generic HVSHELL library have an equivalent in the Haskell Prelude or libraries, the name of the generic function is prefixed with a ‘g’.

4.2 Module `GH.Library.Bounds`

```
gminBound {t} :: (gminBound {t}) ⇒ t
gmaxBound {t} :: (gmaxBound {t}) ⇒ t
gminBound, gmaxBound {t :: *}      :: t
gminBound, gmaxBound {t :: * → *} :: a → t a
```

These are slight generalisations of the *minBound* and *maxBound* members of the `Bounded` type class. They have the property that for all types *t* of kind `*`:

```
∀ a :: t . gminBound {t} ≤ a ≤ gmaxBound {t}
```

However, these functions are also defined for types for which `Bounded` is not derivable; i.e., types which are not enumerations or simple product types [10, Appendix D].

4.3 Module `GH.Library.Collect`

The functions in this module collect information about types, their constructors and their labels.

```
constructorOf {t}          :: (constructorOf {t}) ⇒ t → ConDescr
constructorOf {t :: *}    :: t → ConDescr
constructorOf {t :: * → *} :: (a → ConDescr) → t a → ConDescr
```

The function *constructorOf* returns a description of the topmost constructor in a value.

```
constructors {t}          :: (constructors {t}) ⇒ [ConDescr]
constructors {t :: *}    :: [ConDescr]
constructors {t :: * → *} :: [ConDescr] → [ConDescr]
```

The function *constructors* returns a list of descriptions of all topmost constructors used in a datatype.

```

labelsOf {t}                :: (labelsOf {t}) ⇒ t → [LabelDescr]
labelsOf {t :: *}           :: t → [LabelDescr]
labelsOf {t :: * → *}      :: (a → LabelDescr) → t a → LabelDescr

```

The function *labelsOf* returns a list of descriptions of labels in a value, or the empty list when the current constructor has no labels.

```

labels {t}                  :: (labels {t}) ⇒ [LabelDescr]
labels {t :: *}             :: [LabelDescr]
labels {t :: * → *}        :: [LabelDescr] → [LabelDescr]

```

The function *labels* returns a list of descriptions of labels for a type, or the empty list when the datatype has no constructors with labels.

```

constructorsAndLabels {t}  :: (constructorsAndLabels {t}, labels {t}) ⇒
                               [(ConDescr, [LabelDescr])]
constructorsAndLabels {t :: *} :: [(ConDescr, [LabelDescr])]
constructorsAndLabels {t :: * → *}
                               :: [(ConDescr, [LabelDescr])] → [(ConDescr, [LabelDescr])]

```

The function *constructorsAndLabels* combines the above information: it returns a list of all constructors, paired with the labels present in the given type constructor.

The definitions of *ConDescr* and *LabelDescr* are given in Section 3.2.

4.4 Module `GH.Library.Compare`

```

gcompare {t}               :: (gcompare {t}) ⇒ t → t → Ordering
gcompare {t :: *}          :: t → t → Ordering
gcompare {t :: * → *}      :: (a → a → Ordering) → t a → t a → Ordering

```

The function *gcompare* is the generic version of *compare* in the *Ord* class.

4.5 Module `GH.Library.DeepSeq`

```

dSeq {t | b} :: (dSeq {t | b}) ⇒ t → b → b

```

The function *dSeq* is a variant of the standard function *seq*, which evaluates its first argument completely before returning the second.

This module also defines the type class *DeepSeq*

```

class DeepSeq a where
  deepSeq :: a → b → b

```

and the operator

```
($!) :: (DeepSeq a) => (a -> b) -> a -> b
```

– completely strict application, a variant of (\$!).

Class instances for `DeepSeq` are provided, using the generic function `dSeq`, for a few standard datatypes.

4.6 Module `GH.Library.Enum`

```
empty {t}          :: (empty {t}) => t
empty {t :: *}     :: t
empty {t :: * -> *} :: a -> t a
```

The `empty` function generates a default value of a datatype. For datatypes with multiple constructors, the leftmost constructor is preferred.

```
enum {t}          :: (enum {t}) => [t]
enum {t :: *}     :: [t]
enum {t :: * -> *} :: [a] -> [t a]
```

The function `enum` enumerates all values of a datatype in a (possibly infinite) list. For infinite datatypes, the function applies diagonalisation such that each value appears at a finite position of the list.

4.7 Module `GH.Library.Eq`

```
eq {t}           :: (enum {t}, eq {t}) => t -> t -> Bool
eq {t :: *}      :: t -> t -> Bool
eq {t :: * -> *} :: [a] -> (a -> a -> Bool) -> t a -> t a -> Bool
```

The function `eq` is the generic version of `(==)` in the `Eq` class. It depends on `enum` so that functions with finite domain can be compared.

```
neq {t}          :: (enum {t}, eq {t}) => t -> t -> Bool
neq {t :: *}     :: t -> t -> Bool
```

The function `neq` is the generic version of `(/=)`. It is defined to be the negation of `eq`.

4.8 Module `GH.Library.Map`

```
gmap {t1, t2}    :: (gmap {t1, t2}) => t1 -> t2
gmap {t :: *}    :: t -> t
gmap {t :: * -> *} :: (a -> b) -> t a -> t b
```

The function `gmap` is the generic version of `fmap` in the `Functor` class.

4.9 Module `GH.Library.MapM`

```
mapMl {t1, t2 | m}      :: (mapMl {t1, t2 | m}, Functor m, Monad m) => t1 -> m t2
mapMr {t1, t2 | m}      :: (mapMr {t1, t2 | m}, Functor m, Monad m) => t1 -> m t2
mapMl, mapMr {t :: *}    :: (Functor m, Monad m) => t -> m t
mapMl, mapMr {t :: * -> *} :: (Functor m, Monad m) => (a -> m b) -> t a -> m (t b)
```

These are the generic versions of the monadic map `mapM` in the Prelude. The function `mapMl` traverses a data structure from left to right (just like `mapM`), while `mapMr` traverses from right to left. The `Monad` in the context should also be an instance of class `Functor`.

4.10 Module `GH.Library.ReadShow`

```
gshowsPrec {t} :: (gshowsPrec {t} =>
  Bool -> -- record?
  Int -> -- precedence
  a -> ShowS
greadsPrec {t} :: (greadsPrec {t} =>
  Bool -> -- record?
  Int -> -- precedence
  ReadS a
gshowsPrec {t :: *}      :: Bool -> Int -> t -> ShowS
gshowsPrec {t :: * -> *} :: (Bool -> Int -> a -> ShowS) -> Bool -> Int -> t a -> ShowS
greadsPrec {t :: *}      :: Bool -> Int -> ReadS t
greadsPrec {t :: * -> *} :: (Bool -> Int -> ReadS a) -> Bool -> Int -> ReadS (t a)
```

The generic versions of `show` and `read` (in classes `Show` and `Read`).

The extra argument of type `Bool` is used internally to specify whether field labels are to be printed (and separated by commas). It should usually be `False`.

Since calling these functions is a bit cumbersome, the following specialisations are provided:

```
gshow {t}          :: (gshowsPrec {t}) => t -> String
gshow1 {t | a}     :: (gshowsPrec {t}, Show a) => t a -> String
gread {t}         :: (greadsPrec {t}) => String -> t
gread1 {t | a}    :: (greadsPrec {t}, Read a) => String -> t a
gshow {t :: *}     :: t -> String
gshow1 {t :: * -> *} :: Show a => t a -> String
gread {t :: *}     :: String -> t
gread1 {t :: * -> *} :: Read a => String -> t a
```


4.11 Module `GH.Library.Reduce`

```
rreduce {t | b}      :: (rreduce {t | b}) ⇒ t → b → b
lreduce {t | b}      :: (lreduce {t | b}) ⇒ b → t → b
rreduce {t :: * }      :: t → b → b
rreduce {t :: * → * }  :: (a → b → b) → t a → b → b
lreduce {t :: * }      :: b → t → b
lreduce {t :: * → * }  :: (b → a → b) → b → t a → b
```

The function *rreduce* is a generic version of *foldr* (note the reversed order of the last two arguments!), while *lreduce* is a generic *foldl* ([2, Section 5.4] and [8, Section 12.2]).

```
crush {t | a}        :: (lreduce {t | a}) ⇒ (a → a → a) → a → t a → a
crush {t :: * → * }   :: (a → a → a) → a → t a → a
```

The function *crush* is an instance of *lreduce* with a slightly more familiar type.

The following functions are all defined in terms of the above functions, and most have counterparts in the Haskell Prelude:

```
gsum, gproduct {t | a}      :: (lreduce {t | a}, Num a) ⇒ t a → a
gand, gor {t}                :: (lreduce {t | Bool}) ⇒ t → Bool → Bool
flatten {t | a}              :: (rreduce {t | [a]}) ⇒ t a → [a]
count {t | a}                :: (rreduce {t | Int}) ⇒ t a → Int
comp {t | a}                 :: (lreduce {t | a → a}) ⇒ t (a → a) → (a → a)
gconcat {t | a}              :: (lreduce {t | [a]}) ⇒ t [a] → [a]
gall {t | a}                 :: (lreduce {t | Bool}) ⇒ (a → Bool) → t a → Bool
gany {t | a}                 :: (rreduce {t | Bool}) ⇒ (a → Bool) → t a → Bool
gelem {t | a}                :: (rreduce {t | Bool}, Eq a) ⇒ a → t a → Bool
gsum, gproduct {t :: * → * }  :: Num a ⇒ t a → a
gand, gor {t :: * → * }      :: t Bool → Bool
flatten {t :: * → * }          :: t a → [a]
count {t :: * → * }           :: t a → Int
comp {t :: * → * }            :: t (a → a) → (a → a)
gconcat {t :: * → * }         :: t [a] → [a]
gall, gany {t :: * → * }     :: (a → Bool) → t a → Bool
gelem {t :: * → * }          :: Eq a ⇒ a → t a → Bool
```

The function *flatten* collects all values of type *a* in a list, and *comp* composes all functions contained in a datatype.

4.12 Module `GH.Library.Table`

The module `Table` provides a type-indexed type and functions for building memo tables of functions.

This module has not been updated to Dependency-style and EGH syntax, because type-indexed type support is currently lacking.

4.13 Module `GH.Library.ZipWith`

```
gzipWith {t1, t2, t3} :: (gzipWith {t1, t2, t3} ⇒ (t1, t2) → Maybe t3)  
gzipWith {t :: *} :: (t, t) → Maybe t  
gzipWith {t :: * → *} :: ((a, b) → Maybe c) → (t a, t b) → Maybe (t c)
```

A generic version of *zipWith*, except that it returns a Maybe value, the result being *Nothing* when the two data structures do not have the same shape.

```
gunzipWith {t1, t2, t3} :: (gunzipWith {t1, t2, t3} ⇒ t1 → (t2, t3))  
gunzipWith {t :: *} :: t → (t, t)  
gunzipWith {t :: * → *} :: (a → (b, c)) → t a → (t b, t c)
```

The function *gunzipWith* is a generic version of *unzip*.

```
gzip {t | a, b} :: (gzipWith {t, t, t} ⇒ t a → t b → Maybe (t (a, b)))  
gunzip {t | a, b} :: (gunzipWith {t, t, t} ⇒ t (a, b) → (t a, t b))  
gzip {t :: * → *} :: t a → t b → Maybe (t (a, b))  
gunzip {t :: * → *} :: t (a, b) → (t a, t b)
```

These functions are more or less direct generalisations of *zip* and *unzip* respectively, defined via generic abstraction as instances of *gzipWith* and *gunzipWith*.

5 Future Work

In the future, we plan to continue our work on the compiler. Among the many possible extensions and improvements, we are initially considering:

- improved support for type-indexed data types
- adding a type checker and better support for generic type inference
- a view mechanism (i.e., implicit maps between data types); better support for fixpoints
- ...

As we have not yet decided how the next major release of the `Generic HASKELL` compiler will look, these topics are subject to change. Any input and feedback is most welcome!

6 Meta-information

6.1 Contact

The Generic HVSHELL Project For information regarding the Generic HVSHELL project have a look at <http://generic-haskell.org> or send email to info@generic-haskell.org.

Mailing List A low volume mailing list exists. Currently it serves as a place for distributing information relevant to Generic HVSHELL and for announcing our project meetings. This is the appropriate forum for general language discussions and whatnot. The address is generic-haskell@cs.uu.nl. To subscribe to the mailing list, point your browser to <https://mail.cs.uu.nl/mailman/listinfo/generic-haskell> and follow the instructions.

Bug Reports Bugs can be reported to bugs@generic-haskell.org.

6.2 Caveats

The Generic HVSHELL compiler is a research prototype. Many of its features, especially the more experimental ones, may change as we gain more experience and understanding.

It should be noted that the compiler does not perform type checking of the Generic HVSHELL source language. Thus type errors in Generic HVSHELL source will often be discovered only when the generated Haskell source is compiled.

6.3 Known bugs and limitations

1. The constructor descriptors for user-defined data types that have infix constructors with non-default fixity will be generated incorrectly with the default fixity.
2. The implementation of type-indexed types is lacking in several areas. For example, a type-indexed data type which is specialised to the same type in two separate modules results in types which should be the same, but are treated differently by Haskell. Support for type-indexed types is currently being rewritten and will be updated in the next Generic HVSHELL release.
3. Generic application to type arguments of higher kind does not work for generic abstractions. For example, defining

```
import GH.Library.Eq
myeq {a} :: a -> a -> Bool
myeq {a} = eq {a}
```

you can call `eq {[]} eqa`, but not `myeq {[]} eqa`. However, you can use local redefinition and say

```
let eq {a} = eqa in myeq ([a])
```

4. This list is incomplete.

6.4 Change log

Coral (1.42) Dependency-style Generic Haskell [9] is supported, i.e., generic functions can be written in a simpler and more natural style. In particular, type signatures of generic functions have become simpler – there is no need to define kind-indexed types any longer. Local redefinition, generic abstraction, and default cases are now implemented as described in “Exploring Generic Haskell” [8].

Beryl (1.23) Syntax for using Con and Label in generic functions has slightly changed. Added constructor and default cases. Improved support for the module system. Re-vamped specialisation mechanism – it is now demand driven and generates less code. Numerous bug fixes.

Amber (0.99) The first release.

6.5 Acknowledgements

Thanks to Ralf Hinze for `frown :-()`, to Arthur Baars and Doaitse Swierstra for `ag`, and to Simon Marlow and Sven Panne for the original Happy Haskell grammar.

6.6 Copyright information

`gh` – a compiler for Generic HASKELL.

Copyright © 2001 – 2005 The Generic HASKELL Team. Utrecht University

This library is free software; you can redistribute it and/or modify it under the terms of the GNU Lesser General Public License as published by the Free Software Foundation; either version 2.1 of the License, or (at your option) any later version.

This library is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU Lesser General Public License for more details.

You should have received a copy of the GNU Lesser General Public License along with this library; if not, write to the Free Software Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA 02111-1307 USA

Bibliography

- [1] Dave Clarke and Andres Löb. Generic Haskell, specifically. In J. Gibbons and J. Jeuring, editors, *Generic Programming*, IFIP, pages 21–47. Kluwer Academic Publishers, 2003.
- [2] Ralf Hinze. *Generic Programs and Proofs*. 2000. Habilitationsschrift, Bonn University.
- [3] Ralf Hinze. Polytypic values possess polykinded types. In Roland Backhouse and José Nuno Oliveira, editors, *Mathematics of Program Construction*, volume 1837 of *LNCS*, pages 2–27. Springer-Verlag, 2000.
- [4] Ralf Hinze and Johan Jeuring. Generic Haskell: applications. In *Generic Programming, Advanced Lectures*, volume 2793 of *LNCS*, pages 57–97. Springer-Verlag, 2003.
- [5] Ralf Hinze and Johan Jeuring. Generic Haskell: practice and theory. In *Generic Programming, Advanced Lectures*, volume 2793 of *LNCS*, pages 1–56. Springer-Verlag, 2003.
- [6] Ralf Hinze, Johan Jeuring, and Andres Löb. Type-indexed data types. In *Proceedings of the 6th Mathematics of Program Construction Conference, MPC'02*, volume 2386 of *LNCS*, pages 148–174, 2002.
- [7] Patrik Jansson and Johan Jeuring. PolyP — a polytypic programming language extension. In *Conference Record of POPL '97: The 24th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, pages 470–482. ACM Press, 1997.
- [8] Andres Löb. *Exploring Generic Haskell*. PhD thesis, Utrecht University, September 2004.
- [9] Andres Löb, Dave Clarke, and Johan Jeuring. Dependency-style Generic Haskell. In Olin Shivers, editor, *Proceedings of the International Conference on Functional Programming, ICFP'03*, pages 141–152. ACM Press, August 2003. Also appeared as Technical Report UU-CS-2003-022, Institute of Information and Computing Sciences, Utrecht University.
- [10] Simon Peyton Jones, John Hughes, et al. Haskell 98 — A non-strict, purely functional language. Available from <http://haskell.org>, Feb 1999.