Deriving class instances for datatypes.*

René Thiemann
August 16, 2018

Abstract

We provide a framework for registering automatic methods to derive class instances of datatypes, as it is possible using Haskell’s “deriving Ord, Show, . . . ” feature.

We further implemented such automatic methods to derive (linear) orders or hash-functions which are required in the Isabelle Collection Framework [1] and the Container Framework [2]. Moreover, for the tactic of Huffman and Krauss to show that a datatype is countable, we implemented a wrapper so that this tactic becomes accessible in our framework.

Our formalization was performed as part of the IsaFoR/CeTA project1 [3]. With our new tactic we could completely remove tedious proofs for linear orders of two datatypes.

Contents

1 Important Information 2

2 Generating linear orders for datatypes 2
   2.1 Introduction .............................................. 2
   2.2 Implementation Notes ................................. 3
   2.3 Features and Limitations ............................. 3
   2.4 Installing the generator ............................... 3

3 Hash functions 4
   3.1 Introduction ........................................... 4
   3.2 Features and Limitations ............................. 4
   3.3 Installing the generator ............................... 4

4 Loading derive-commands 5

*Supported by FWF (Austrian Science Fund) project P22767-N13.
1http://cl-informatikuibk.ac.at/software/ceTA
5 Examples
5.1 Register standard existing types ........................................ 5
5.2 Without nested recursion .................................................. 5
5.3 Using other datatypes .......................................................... 5
5.4 Explicit mutual recursion .................................................... 6
5.5 Implicit mutual recursion .................................................... 6
5.6 Examples from IsaFoR ......................................................... 6
5.7 A complex datatype .............................................................. 6
6 Acknowledgements ................................................................. 7

1 Important Information

The described generators are outdated as they are based on the old datatype package. Generators for the new datatypes are available in the AFP entry “Deriving”.

theory Derive-Aux
imports Deriving.Derive-Manager
begin

ML-file derive-aux.ML

end

2 Generating linear orders for datatypes

theory Order-Generator
imports Derive-Aux
begin

2.1 Introduction

The order generator registers itself at the derive-manager for the classes ord, order, and linorder. To be more precise, it automatically generates the two functions (≤) and (<) for some datatype dtype and proves the following instantiations.

- instantiation dtype :: (ord,...,ord) ord
- instantiation dtype :: (order,...,order) order
- instantiation dtype :: (linorder,...,linorder) linorder
All the non-recursive types that are used in the datatype must have similar instantiations. For recursive type-dependencies this is automatically generated.

For example, for the datatype \texttt{tree = Leaf nat | Node "tree list"} we require that \texttt{nat} is already in \texttt{linorder}, whereas for \texttt{list} nothing is required, since for the \texttt{tree} datatype the \texttt{list} is only used recursively.

However, if we define \texttt{datatype tree = Leaf "nat list" | Node tree tree} then \texttt{list} must provide the above instantiations.

Note that when calling the generator for \texttt{linorder}, it will automatically also derive the instantiations for \texttt{order}, which in turn invokes the generator for \texttt{ord}. A later invocation of \texttt{linorder} after \texttt{order} or \texttt{ord} is not possible.

### 2.2 Implementation Notes

The generator uses the recursors from the datatype package to define a lexicographic order. E.g., for a declaration \texttt{datatype 'a tree = Empty | Node 'a tree 'a 'a tree} this will semantically result in

\[
\begin{align*}
(\text{Empty} < \text{Node } _ _ _) &= \text{True} \\
(\text{Node } l1 l2 l3 < \text{Node } r1 r2 r3) &= (l1 < r1 || l1 = r1 && (l2 < r2 || l2 = r2 && l3 < r3)) \\
(_ < _) &= \text{False} \\
(l <= r) &= (l < r || l = r)
\end{align*}
\]

The desired properties (like \([x < y; y < z] \implies x < z\)) of the orders are all proven using induction (with the induction theorem from the datatype on \(x\)), and afterwards there is a case distinction on the remaining variables, i.e., here \(y\) and \(z\). If the constructors of \(x\), \(y\), and \(z\) are different always some basic tactic is invoked. In the other case (identical constructors) for each property a dedicated tactic was designed.

### 2.3 Features and Limitations

The order generator has been developed mainly for datatypes without explicit mutual recursion. For mutual recursive datatypes—like \texttt{datatype a = C b and b = D a a}—only for the first mentioned datatype—here \texttt{a}—the instantiations of the order-classes are derived.

Indirect recursion like in \texttt{datatype tree = Leaf nat | Node "tree list"} should work without problems.

### 2.4 Installing the generator

\texttt{lemma linear-cases: (x :: 'a :: linorder) = y \lor x < y \lor y < x by auto}

\texttt{ML-file order-generator.ML}
3 Hash functions

theory Hash-Generator
imports Collections.HashCode Derive-Aux
begin

3.1 Introduction

The interface for hash-functions is defined in the class hashable which has been developed as part of the Isabelle Collection Framework \cite{1}. It requires a hash-function (hashcode), a bounded hash-function (bounded-hashcode), and a default hash-table size (def-hashmap-size).

The hashcode function for each datatype are created by instantiating the recursors of that datatype appropriately. E.g., for \texttt{datatype 'a test = C1 'a 'a | C2 'a test list} we get a hash-function which is equivalent to

\[
\text{hashcode (C1 a b)} = c1 \ast \text{hashcode a} + c2 \ast \text{hashcode b} \\
\text{hashcode (C2 Nil)} = c3 \\
\text{hashcode (C2 (a # as))} = c4 \ast \text{hashcode a} + c5 \ast \text{hashcode as}
\]

where each \(c_i\) is a non-negative 32-bit number which is dependent on the datatype name, the constructor name, and the occurrence of the argument (i.e., in the example \(c1\) and \(c2\) will usually be different numbers.) These parameters are used in linear combination with prime numbers to hopefully get some useful hash-function.

The bounded-hashcode functions are constructed in the same way, except that after each arithmetic operation a modulo operation is performed.

Finally, the default hash-table size is just set to 10, following Java’s default hash-table constructor.

3.2 Features and Limitations

We get same limitation as for the order generator. For mutual recursive datatypes, only for the first mentioned datatype the instantiations of the hashable-class are derived.

3.3 Installing the generator

\texttt{lemma hash-mod-lemma: 1 < (n :: nat) \implies x mod n < n by auto}

\texttt{ML-file hash-generator.ML}
4 Loading derive-commands

theory Derive
imports
  Order-Generator
  Hash-Generator
  Deriving.Countable-Generator
begin
  We just load the commands to derive (linear) orders, hash-functions, and
  the command to show that a datatype is countable, so that now all of them
  are available. There are further generators available in the AFP entries of
  lightweight containers and Show.
  print-derives
end

5 Examples

theory Derive-Examples
imports
  Derive
  HOL.Rat
begin

5.1 Register standard existing types
derive linorder list sum prod

5.2 Without nested recursion
datatype 'a bintree = BEmpty | BNode 'a bintree 'a 'a bintree
derive linorder bintree
derive hashable bintree
derive countable bintree

5.3 Using other datatypes
datatype nat-list-list = NNil | CCons nat list nat-list-list
derive linorder nat-list-list
derive hashable nat-list-list
derive countable nat-list-list
5.4 Explicit mutual recursion

```
datatype 'a mtree = MEmpty | MNode 'a 'a mtree-list and
'a mtree-list = MNil | MCons 'a mtree 'a mtree-list
```

derive linorder mtree
derive hashable mtree
derive countable mtree

5.5 Implicit mutual recursion

```
datatype 'a tree = Empty | Node 'a 'a tree list
datatype-compat tree
derive linorder tree
derive hashable tree
derive countable tree
datatype 'a ttree = TEmpty | TNode 'a 'a ttree list tree
datatype-compat ttree
derive linorder ttree
derive hashable ttree
derive countable ttree
```

5.6 Examples from IsaFoR

```
datatype ('f,'v) term = Var 'v | Fun 'f ('f,'v) term list
datatype-compat term
datatype ('f, 'l) lab =
  Lab ('f, 'l) lab 'l
| FunLab ('f, 'l) lab ('f, 'l) lab list
| UnLab 'f
| Sharp ('f, 'l) lab
datatype-compat lab
derive linorder term lab
derive countable term lab
derive hashable term lab
```

5.7 A complex datatype

The following datatype has nested indirect recursion, mutual recursion and uses other datatypes.
\textbf{datatype} \((\ 'a, 'b)\) complex = \\
\hspace{1em}C1 \ 'a \ tree \\
\hspace{1em}C2 \ ('a, 'b) \ complex \ list \ tree \ tree \ 'b \ ('a, 'b) \ complex \ ('a, 'b) \ complex2 \ tree \ list \\
\textbf{and} \ ('a, 'b) \ complex2 = D1 \ ('a, 'b) \ complex \ tree

\textbf{datatype-compat} complex complex2

\textbf{derive} linorder complex
\textbf{derive} hashable complex
\textbf{derive} countable complex

\textbf{end}

6 Acknowledgements

We thank

- Lukas Bulwahn and Brian Huffman for the discussion on a generic derive command and the pointer to the tactic for countability.

- Alexander Krauss for pointing me to the recursors of the datatype package.

- Peter Lammich for the inspiration of developing a hash-function generator.

- Andreas Lochbihler for the inspiration of developing generators for the container framework.

- Christian Urban for his cookbook about the ML-level of Isabelle.

- Stefan Berghofer, Cezary Kaliszyk, and Tobias Nipkow for their explanations on several Isabelle related questions.

References

