

Formal foundations for GADTs in Scala

Radosław Waśko

9 maja 2020

- 1 Czym są GADT?
- 2 Enkodowanie GADT w Scali
- 3 pDOT
- 4 Enkodowanie GADT w pDOT

Czym są GADT? Przykłady

GADT pozwalają nam definiować dodatkowe więzy na typach zawartych w strukturze.

Na przykład - otypowane drzewa wyrażeń

```
{-# LANGUAGE GADTs #-}
```

```
data Expr a where
```

```
  Lit  :: Int  -> Expr Int
```

```
  Plus :: Expr Int -> Expr Int -> Expr Int
```

```
  Pair :: Expr b -> Expr c -> Expr (b, c)
```

```
eval :: Expr a -> a
```

```
eval (Lit x) = x -- here a ::= Int
```

```
eval (Plus lhs rhs) = (eval lhs) + (eval rhs)
```

```
eval (Pair lhs rhs) = (eval lhs, eval rhs) -- a ::= (b, c)
```

```
eval (Pair (Plus (Lit 40) (Lit 2)) (Lit 23)) // (42, 23)
```

GADT w Scali - przykłady

GADT definiujemy tak jak ADT dodając więzy typowe poprzez klauzulę `extends`.

W Scali 3 (dotty):

```
enum Expr[A] {
  case Lit(x: Int) extends Expr[Int] // A ::= Int
  case Plus(lhs: Expr[Int], rhs: Expr[Int]) extends Expr[Int]
  case Pair[B,C](lhs: Expr[B], rhs: Expr[C]) extends Expr[(B, C)]
}
```

W Scali 2:

```
sealed trait Expr[A] {}
case class Lit(x: Int) extends Expr[Int] // A ::= Int
case class Plus(lhs: Expr[Int], rhs: Expr[Int]) extends Expr[Int]
case class Pair[B,C](lhs: Expr[B], rhs: Expr[C]) extends Expr[(B, C)]
```

GADT w Scali - przykłady

```
enum Expr[A] {
  case Lit(x: Int) extends Expr[Int]
  case Plus(lhs: Expr[Int], rhs: Expr[Int]) extends Expr[Int]
  case Pair[B,C](lhs: Expr[B], rhs: Expr[C]) extends Expr[(B, C)]
}
```

```
def eval[A](e: Expr[A]): A = e match {
  case Lit(x) => x // A ::= Int
  case Plus(lhs, rhs) => eval(lhs) + eval(rhs) // A ::= Int
  case Pair(lhs, rhs) => (eval(lhs), eval(rhs)) // A ::= (B, C)
}
```

```
eval(Pair(Plus(Lit(40), Lit(2)), Lit(23))) // (42, 23)
```

Definicja GADT

```
enum Expr[A] {
  case Lit(x: Int) extends Expr[Int] // A == Int
  case Plus(lhs: Expr[Int], rhs: Expr[Int]) extends Expr[Int]
  case Pair[B,C](lhs: Expr[B], rhs: Expr[C]) extends Expr[(B, C)]
}
```

GADT cechują się:

- możliwością definiowania więzów na typach
- rekurencyjnymi odwołaniami
- typami egzystencjalnymi

Definicja GADT

Zatem GADT możemy zdefiniować w rachunku z typami egzystencjalnymi i rekurencyjnymi oraz więzami typowymi. Ogólny typ GADT według Xi et al.¹:

$$T \equiv \mu t. \lambda \bar{\alpha}. (\exists [\bar{\alpha}_1, \bar{\sigma}_1 = \bar{\alpha}]. \tau_1 + \cdots + \exists [\bar{\alpha}_n, \bar{\sigma}_n = \bar{\alpha}]. \tau_n)$$

¹w *Guarded recursive datatype constructors* [1]

Definicja GADT

Łatwiej będzie tę ogólną definicję zrozumieć na przykładzie:
Weźmy poprzedni przykład:

```
enum Expr[A] {  
  case Lit(x: Int) extends Expr[Int]  
  case Plus(lhs: Expr[Int], rhs: Expr[Int]) extends Expr[Int]  
  case Pair[B,C](lhs: Expr[B], rhs: Expr[C]) extends Expr[(B, C)]  
}
```

$$\begin{aligned} Expr \equiv \mu t. \lambda A. (&\exists[A = Int]. Int \\ &+ \exists[A = Int]. t\ A * t\ A \\ &+ \exists[B, C, A = B * C]. t\ B * t\ C) \end{aligned}$$

Enkodowanie GADT

Scala od dawna wspiera typy egzystencjalne i rekurencyjne.
Spróbujmy zakodować więzy typów.

Enkodowanie więzów - funkcje

Pierwszym przybliżeniem kodowania więzów mogą być funkcje²:

```
sealed trait Eq[A,B] {  
  def to(x: A): B  
  def from(y: B): A  
}
```

```
case class Refl[A]() extends Eq[A,A] {  
  def to(x: A): A = x  
  def from(y: A): A = y  
}
```

²Pomysł oparty na pracy *GADTs for the OCaml masses* [2]

Enkodowanie więzów - funkcje

Instancja klasy `Eq` służy nam za świadka, że typy `A` i `B` są równe. Każdy więz typowy zostaje przekształcony w dodatkowego członka danego wariantu klasy.

```
sealed trait Expr[A]
case class LitC[A](x: Int, ev: Eq[A, Int]) extends Expr[A]
case class PlusC[A](lhs: Expr[Int], rhs: Expr[Int], ev: Eq[A, Int])
  extends Expr[A]
case class PairC[A,B,C](lhs: Expr[B], rhs: Expr[C], ev: Eq[A, (B,C)])
  extends Expr[A]
def Lit(x: Int): Expr[Int] = LitC(x, Refl())
def Plus(lhs: Expr[Int], rhs: Expr[Int]): Expr[Int] =
  PlusC(lhs, rhs, Refl())
def Pair[B,C](lhs: Expr[B], rhs: Expr[C]): Expr[(B,C)] =
  PairC(lhs, rhs, Refl())
```

W konstruktorze równości przywołujemy za pomocą `Refl`. Ponadto jeśli zabronimy tworzenia innych instancji `Eq`, możemy zagwarantować, że odpowiednie instancje `Expr` powstały tylko z niesprzecznych równości.

Enkodowanie więzów - funkcje

```
trait Eq[A,B] {  
  def to(x: A): B  
  def from(y: B): A  
}  
  
def eval[A](e: Expr[A]): A = e match {  
  case LitC(x, ev) => ev.from(x)  
  case PlusC(lhs, rhs, ev) => ev.from(eval(lhs) + eval(rhs))  
  case PairC(lhs, rhs, ev) => ev.from((eval(lhs), eval(rhs)))  
}
```

W miejscu gdzie korzystamy z poszczególnych równości, wstawiamy wywołania odpowiednich funkcji klasy Eq.

Enkodowanie więzów - funkcje

Jedną z wad tego kodowania jest jego słaba integracja z innymi elementami systemu typów, jak np. ze strukturami kowariantnymi.

```
def lista[A,B](lst: List[A], ev: Eq[A, B]): List[B] =  
  lst.map(ev.to)
```

Ta metoda realizuje równość nie bezpośrednio w systemie typów, ale niejako tworząc konwersje pomiędzy niezwiązanymi typami, które muszą zostać aplikowane do każdego elementu, na który wpływa dana równość.

W szczególności jak widać na powyższym przykładzie, konwertowanie bardziej złożonych struktur może wiązać się z niepotrzebnym przechodzeniem po całej (potencjalnie dużej) strukturze.

Enkodowanie więzów - type members

Kolejne podejście³ jest podobne do poprzedniego, ale działa już bliżej samego systemu typów, a nie runtime'u.

```
trait Eq[A,B] {  
  type To >: B <: A  
  type From >: A <: B  
}  
  
case class Refl[A]() extends Eq[A,A] {  
  type To = A  
  type From = A  
}
```

³Bazuje na *Towards improved GADT reasoning in Scala* [3]

Enkodowanie więzów - type members

Korzystamy tu z faktu, że konkretne instancje klasy mogą zawierać tylko członki typowe o niesprzecznych więzach.

```
trait Contradiction {  
  type C >: String <: Int  
}
```

```
object X extends Contradiction // Error  
// object X cannot be instantiated since  
// it has a member C with possibly conflicting bounds  
// String <: ... <: Int
```

Enkodowanie więzów - type members

Konstruktory pozostają dokładnie takie same jak w poprzednim podejściu. Zmienia się sposób w jaki korzystamy ze świadków.

```
def eval[A](e: Expr[A]): A = e match {
  case LitC(x, ev) => x: ev.T
  case PlusC(lhs, rhs, ev) => eval(lhs) + eval(rhs): ev.T
  case PairC(lhs, rhs, ev) => (eval(lhs), eval(rhs)): ev.T
}
```

Dodając odpowiednią adnotację typową (nie jest to type-cast) dajemy wskazówkę kompilatorowi. `ev` jest typu `Eq[A, Int]`, zatem `ev.T >: Int <: A`. Kompilator widzi, że skoro `x: Int`, to możemy do niego przypisać typ `ev.T`, który jest nadtypem `Int`, zaś skoro `ev.T` jest podtypem `A`, to możemy użyć wartości `x` tam gdzie oczekujemy wartości typu `A`.

Enkodowanie więzów - type members

To kodowanie jest dobrym punktem wyjścia, w szczególności lepiej radzi sobie z nakreślonym wcześniej problemem złożonych struktur:

```
def lista[A,B](lst: List[A], ev: Eq[A, B]): List[B] =  
  lst: List[ev.F]
```

Ta adnotacja typowa jest tylko wskazówką dla kompilatora, nic nas nie kosztuje w runtime.

pDOT⁴ (Path-dependent Object Types) jest rachunkiem lambda formalizującym podstawy, na których opiera się system typów Scali 3.

Jego najważniejsze cechy:

- type members
- intersection types
- path-dependent types
- singleton types

pozwalają na zakodowanie takich funkcjonalności jak generic classes czy parametric polymorphism.

⁴Opisany w *A path to DOT* [4]

pDOT - Type members

```

trait T { type A >: Nothing <: Any }
trait U { type A >: L <: H }
  
```

$$T = \{A : \perp..T\}$$

$$U = \{A : L..H\}$$

$$\frac{\Gamma \vdash p : \{A : L..H\}}{\Gamma \vdash p.A <: H}$$

pDOT - Singleton types

```
val x = 5
def f(v: x.type) = ...
f(x)
f(5) // Type Mismatch

      let x = 5 in
      let f =  $\lambda(v : x.type)$  ... in
      f x
```

pDOT - Self-type (typy rekurencyjne)

```
trait T { self =>  
  val x: self.type  
}
```

```
val t = new T { val x = this }
```

$$T = \mu(\text{self} : \{x : \text{self.type}\})$$
$$t = \nu(\text{self} : T)\{x = \text{self}\}$$

pDOT - Cykliczne odwołania

Umieszczamy definicje wewnątrz modułu

```
object Module {
  def f(x: Int): Int = g(x)
  def g(x: Int): Int = f(x)
}
Module = ν(module : μ(module : {
  f : ∀(x : Int) Int; g : ∀(x : Int) Int
})){f = λ(x : Int) module.g x
  g = λ(x : Int) module.f x}
```

pDOT - Intersection types

```

trait A { def g(v: Int): String }
trait B { val x: Int }
def f(obj: A & B): String = obj.g(obj.x)

```

```

class AB extends A with B {
  def g(v: Int): String = v.toString
  val x: Int = 42
}
f(new AB)

```

$$A = \{g : \forall (v : \text{Int}) \text{String}\}$$

$$B = \{x : \text{Int}\}$$

$$f : \forall (obj : A \wedge B) \text{String}$$

$$f = \lambda (obj : A \wedge B) \text{obj.g } obj.x$$

$$\text{let } o = v(\text{self} : A \wedge B)\{g = \dots; x = 42\} \text{ in } f \ o$$

pDOT - Path-dependent types

```
trait T {  
  type A  
  val x: A  
}
```

```
def f(v: T): v.A = v.x
```

$$T = \mu(\text{self} : \{A : \perp..T\} \wedge \{x : \text{self}.A\})$$

$$f : \forall(v : T) v.A$$

$$f = \lambda(v : T) v.x$$

pDOT - przykład: jak zakodować polimorficzną funkcję

```
def id[A](x: A): A = x
id[Int](42)
```

```
trait TypeLabel { type Typ }
def id(T: TypeLabel)(x: T.Typ): T.Typ = x
id(new TypeLabel { type Typ = Int })(42)
```

$$\begin{aligned}
 \text{TypeLabel} &= \{ \text{Typ} : \perp..T \} \\
 \text{id} &: \forall (T : \text{TypeLabel}) \forall (x : T.\text{Typ}) \ T.\text{Typ} \\
 \text{id} &= \lambda (T : \text{TypeLabel}) \lambda (x : T.\text{Typ}) \ x \\
 \text{id} \ (\nu(_ : \text{TypeLabel}) \{ \text{Typ} = \text{Int} \}) & \ 42
 \end{aligned}$$

pDOT - przykład: lista

```

ν(sci ⇒ List = μ(self: {A: ⊥..⊤} ∧ {head: ∀(⊔) self.A} ∧ {tail: ∀(⊔) (sci.List ∧ {A: ⊥..self.A)}});
  nil: ∀(x: {A: ⊥..⊤}) sci.List ∧ {A: ⊥..⊥}
    = λ(x: {A: ⊥..⊤}) let result = ν(self ⇒ A = ⊥;
                                     head: ∀(y: ⊤) self.A = λ(y: ⊤) self.head y;
                                     tail: ∀(y: ⊤) (sci.List ∧ self.A) = λ(y: ⊤) self.tail y)
      in result;
  cons: ∀(x: {A: ⊥..⊤}) ∀(hd: x.A) ∀(tl: sci.List ∧ {A: ⊥..x.A}) sci.List ∧ {A: ⊥..x.A}
    = λ(x: {A: ⊥..⊤}) λ(hd: x.A) λ(tl: sci.List ∧ {A: ⊥..x.A})
      let result = ν(self ⇒ A = x.A;
                    head: ∀(⊔) self.A = λ_.hd
                    tail: ∀(⊔) (sci.List ∧ self.A) = λ_.tl)
      in result)

```

Przykład zaczerpnięty z [4]

Enkodowanie ADT w pDOT

Zanim podejmiemy się zakodowania GADT, warto zastanowić się jak możemy kodować zwykłe ADT i pattern matching na nich w rachunku lambda z typami.

Dwa najpopularniejsze podejścia to:

1 Kodowanie Scotta

$$\text{List} \equiv \mu t. \forall \alpha. (\text{Int} \rightarrow t \rightarrow \alpha) \rightarrow \alpha \rightarrow \alpha$$

2 Kodowanie Böhma-Berarducciego

$$\text{List} \equiv \forall \alpha. (\text{Int} \rightarrow \alpha \rightarrow \alpha) \rightarrow \alpha \rightarrow \alpha$$

W obu tych kodowaniach ADT jest reprezentowany przez funkcję, która przyjmuje funkcje określające jak przekształcić poszczególne case'y.

Enkodowanie ADT - kodowanie Scotta

$$\text{List} \equiv \mu t. \forall \alpha. (\text{Int} \rightarrow t \rightarrow \alpha) \rightarrow \alpha \rightarrow \alpha$$

```

sum = λlist: List.
      list
      (λhead: Int. λtail: List. head + sum tail)
      (0)

tail = λlist: List.
       list
       (λhead: Int. λtail: List. tail)
       (nil)

```

Enkodowanie ADT - kodowanie Böhma-Berarducci

$$\text{List} \equiv \mu t. \forall \alpha. (\text{Int} \rightarrow t \rightarrow \alpha) \rightarrow \alpha \rightarrow \alpha$$

```
sum = λlist: ∀α. (Int → α → α) → α → α.
```

```
  list
```

```
    (λelem: Int. λaccumulator: Int.
      elem + accumulator)
```

```
  0
```

```
tail = λlist: ∀α. (Int → α → α) → α → α.
```

```
  fst (list
```

```
    (λh: Int. λacc: list * list.
```

```
      (snd acc, cons e (snd acc)))
```

```
    (nil, nil)
```

```
  )
```

Enkodowanie GADT w pDOT - typy egzystencjalne

Typy egzystencjalne również można zakodować za pomocą type memberów.

Weźmy typ $T = \exists \alpha. (\text{Int} \rightarrow \alpha) * (\alpha \rightarrow \text{String})$.

```
trait T {
  type A
  def f: Int => A
  def g: A => String
}
```

```
val x: T = ...
val str: String = x.g(x.f(42))
```

$$T = \mu(\text{self} : \{A : \perp..T; f : \forall(_ : \text{Int}) \text{self}.A; g : \forall(_ : \text{self}.A) \text{String}\})$$

Enkodowanie GADT w pDOT - więzy

W zakodowaniu więzów pomoże nam wykorzystanie intersection types i singleton types.

```
trait Expr {
  type A
  val x: A
}
trait Lit extends Expr {
  type A = Int
}
trait Pair extends Expr {
  type B; type C
  type A = (B, C)
}
```

$$Expr = \mu(\text{self} : \{A : \perp..T; x : \text{self}.A\})$$

$$Lit = Expr \wedge \{A : \text{Int}..Int\}$$

$$Pair = \mu(s : Expr \wedge \{B : \perp..T; C : \perp..T; A : (s.B * s.C) .. (s.B * s.C)\})$$

Weźmy $e : Lit$, wtedy $e.x : e.A$, ale mamy, $e : \{A : \text{Int}..Int\}$, zatem $e.A <: \text{Int}$, więc $e.x : \text{Int}$.

Enkodowanie GADT w pDOT - więzy w pattern matchu

```

trait TL { type T }
trait Expr { self =>
  type A
  def visit(T: TL)(lit: self.type & Lit => T.T): T.T
}
trait Lit extends Expr { self =>
  type A = Int
  val x: Int
  def visit(T: TL)(lit: self.type & Lit => T.T): T.T =
    lit(self)
}
def eval(R: TL)(e: Expr & { type A = R.T }): R.T =
  e.visit(R)(
    (lit: e.type & Lit) => lit.x
    // lit.x: Int := lit.A := e.A := R.T
  )

```


Przykład: zakodowanie Expr

```

trait TL { type T }
sealed trait Expr { self =>
  type A
  def visit(R: TL)
    (lit: LitC & self.type => R.T,
     plus: PlusC & self.type => R.T,
     pair: PairC & self.type => R.T): R.T
}
trait LitC extends Expr { self =>
  type A = Int
  def x: Int
  def visit(R: TL)
    (lit: LitC & self.type => R.T,
     plus: PlusC & self.type => R.T,
     pair: PairC & self.type => R.T): R.T =
    lit(this)
}
trait PlusC extends Expr { self =>
  type A = Int
  def lhs: Expr & { type A = Int }
  def rhs: Expr & { type A = Int }
  def visit(R: TL)(...): R.T = plus(this)
}
trait PairC extends Expr { self =>
  type B
  type C
  type A = (B,C)
  def lhs: Expr & { type A = B }
  def rhs: Expr & { type A = C }
  def visit(R: TL)(...): R.T = pair(this)
}

```

```

def Lit(_x: Int): Expr & { type A = Int } =
  new LitC {
    def x = _x
  }
// other constructors analogous

def eval(R: TL)(e: Expr & { type A = R.T }): R.T =
  e.visit(R)(
    lit => lit.x : lit.A,
    plus =>
      val T = new TL { type T = Int }
      eval(T)(plus.lhs) + eval(T)(plus.rhs): plus.A,
    pair =>
      val TB = new TL { type T = pair.B }
      val TC = new TL { type T = pair.C }
      (eval(TB)(pair.lhs),
       eval(TC)(pair.rhs)) : pair.A
  )

```

Ogólna postać GADT w pDOT

GADT postaci

$$T \equiv \mu t. \lambda \bar{\alpha}. \left(\exists [\bar{\beta}_1, \bar{\sigma}_1 = \bar{\alpha}]. \tau_1 + \cdots + \exists [\bar{\beta}_n, \bar{\sigma}_n = \bar{\alpha}]. \tau_n \right)$$

możemy zakodować następująco:

```
T = μ(s: {
  α1: ⊥...T; ... αm: ⊥...T
  pmatch: ∀(r: {R:⊥...T})
    ∀(c1: ∀(arg: env.TC1 ∧ s.type) r.R)
    ...
    ∀(cn: ∀(arg: env.TCn ∧ s.type) r.R)
    r.R
})
```

Ogólna postać GADT w pDOT

$$T \equiv \mu t. \lambda \bar{\alpha}. \left(\exists [\bar{\beta}_1, \bar{\sigma}_1 = \bar{\alpha}]. \tau_1 + \dots + \exists [\bar{\beta}_n, \bar{\sigma}_n = \bar{\alpha}]. \tau_n \right)$$

Poszczególne warianty kodujemy jako:

$$\begin{aligned} TC_i = \mu (s: \text{env}.T \wedge \{ \\ & \beta_{i,1}: \perp \dots \top; \dots; \beta_{i,m_i}: \perp \dots \top \\ & \alpha_1 = \sigma_{i,1}; \dots; \alpha_m = \sigma_{i,m} \\ & \text{data}: \tau_i \\ \}) \end{aligned}$$

zaś ich konstruktory:

$$\begin{aligned} c_i: \forall (\text{types}: \{\beta_{i,1}: \perp \dots \top; \dots \beta_{i,m_i}: \perp \dots \top\}) \forall (v: \tau_i) \\ \text{env}.TC_i \wedge \{\beta_{i,1} = \text{types}.\beta_{i,1}; \dots; \beta_{i,m_i} = \text{types}.\beta_{i,m_i}\} \end{aligned}$$

Future work

Do tej pory zdefiniowałem reguły zakodowania podzbioru rachunku $\lambda_{2,G\mu}$ (opisanego w [1]) do pDOT.

Dalsze plany:

- 1 Sformalizować kodowanie $\lambda_{2,G\mu}$ do pDOT
- 2 Udowodnić, że zakodowane termy się typują
- 3 Udowodnić, że kodowanie zachowuje semantykę
- 4 Opisać lub pozbyć się uproszczeń oryginalnego rachunku
- 5 Przeanalizować relację GADT z podtypianiem

Bibliografia



Hongwei Xi, Chiyan Chen, and Gang Chen.

Guarded recursive datatype constructors.

SIGPLAN Not., 38(1):224–235, January 2003.



Yitzhak Mandelbaum and Aaron Stump.

Gadts for the ocaml masses.

01 2009.



Lionel Parreaux, Aleksander Boruch-Gruszecki, and Paolo G. Giarrusso.

Towards improved gadt reasoning in scala.

In *Proceedings of the Tenth ACM SIGPLAN Symposium on Scala*, Scala '19, pages 12–16, New York, NY, USA, 2019. ACM.



Marianna Rapoport and Ondřej Lhoták.

A path to dot: Formalizing fully path-dependent types.

Proc. ACM Program. Lang., 3(OOPSLA):145:1–145:29, October 2019.

TODO prezentacja i kod źródłowy dostępne na:
TODO

Dziękuję za uwagę :)