Lecture 22 – Algebraic Data Types (2)

COSE212: Programming Languages

Jihyeok Park



2023 Fall

Recall



- A way to define new types by combining existing types:
 - product type
 - union type
 - sum type (tagged union type)
 - algebraic data type (recursive sum type of product types)
- ATFAE TRFAE with ADTs and pattern matching.
 - Interpreter and Natural Semantics
 - Type Checker and Typing Rules

Recall



- A way to define new types by combining existing types:
 - product type
 - union type
 - sum type (tagged union type)
 - algebraic data type (recursive sum type of product types)
- ATFAE TRFAE with ADTs and pattern matching.
 - Interpreter and Natural Semantics
 - Type Checker and Typing Rules
- Minor changes in the previous lecture:
 - TAFAE to ATFAE
 - type variables to type names

Recall



- A way to define new types by combining existing types:
 - product type
 - union type
 - sum type (tagged union type)
 - algebraic data type (recursive sum type of product types)
- ATFAE TRFAE with ADTs and pattern matching.
 - Interpreter and Natural Semantics
 - Type Checker and Typing Rules
- Minor changes in the previous lecture:
 - TAFAE to ATFAE
 - type variables to type names
- In this lecture, we will discuss on Type Checker and Typing Rules.

Recall: Natural Semantics of ATFAE



```
/* ATFAE */
enum Tree {
  case Leaf(Number)
  case Node(Tree, Number, Tree)
}
Leaf(42) match {
  case Leaf(v) => v
  case Node(1, v, r) => v
}
```

The natural semantics of ATFAE ignores all the types.

Leaf and Node are not types but variant names.

Leaf and Node are **constructors** that take lists of values and produce **variant values** by adding their variant names as tags.

A pattern matching expression takes a variant value and finds the first match case whose name is equal to the variant name of the value.

Contents



1. Type Checker and Typing Rules

Type Environment for ADTs
Well-Formedness of Types
(Recursive) Function Definition and Application
Algebraic Data Types
Pattern Matching

2. Type Soundness of ATFAE

Recall: Type Soundness Algebraic Data Types - Revised (1)

Algebraic Data Types - Nevised (1)

Algebraic Data Types - Revised (2)

Contents



1. Type Checker and Typing Rules

Type Environment for ADTs
Well-Formedness of Types
(Recursive) Function Definition and Application
Algebraic Data Types
Pattern Matching

2. Type Soundness of ATFAE

Recall: Type Soundness

Algebraic Data Types - Revised (1)

Algebraic Data Types - Revised (2)

Type Checker and Typing Rules



Let's **1** design **typing rules** of ATFAE to define when an expression is well-typed in the form of:

$$\Gamma \vdash e : \tau$$

and 2 implement a type checker in Scala according to typing rules:

```
def typeCheck(expr: Expr, tenv: TypeEnv): Type = ???
```

The type checker returns the **type** of e if it is well-typed, or rejects it and throws a **type error** otherwise.

Similar to TRFAE, we will keep track of the **variable types** using a **type environment** Γ as a mapping from variable names to their types.

Type Environments
$$\Gamma \in \mathbb{X} \xrightarrow{\text{fin}} \mathbb{T}$$
 (TypeEnv)

```
type TypeEnv = Map[String, Type]
```



However, we need additional information in type environments about new types defined by **algebraic data types** (ADTs).

Type Environments
$$\in (\mathbb{X} \xrightarrow{\text{fin}} \mathbb{T}) \times (\mathbb{X}_t \xrightarrow{\text{fin}} (\mathbb{X} \xrightarrow{\text{fin}} \mathbb{T}^*))$$
 (TypeEnv)
$$\Gamma(t) = x_1(\tau_{1,1}, \dots, \tau_{1,m_1}) + \dots + x_n(\tau_{n,1}, \dots, \tau_{n,m_n})$$



However, we need additional information in type environments about new types defined by **algebraic data types** (ADTs).

Type Environments
$$\in (\mathbb{X} \xrightarrow{\text{fin}} \mathbb{T}) \times (\mathbb{X}_t \xrightarrow{\text{fin}} (\mathbb{X} \xrightarrow{\text{fin}} \mathbb{T}^*))$$
 (TypeEnv)
$$\Gamma(t) = x_1(\tau_{1,1}, \dots, \tau_{1,m_1}) + \dots + x_n(\tau_{n,1}, \dots, \tau_{n,m_n})$$

and sum types are commutative:

$$\Gamma(\mathtt{A}) = \mathtt{B}(\mathtt{bool}) + \mathtt{C}(\mathtt{num}) \qquad \mathsf{equivalent} \ \mathsf{to} \qquad \Gamma(\mathtt{A}) = \mathtt{C}(\mathtt{num}) + \mathtt{B}(\mathtt{bool})$$



However, we need additional information in type environments about new types defined by **algebraic data types** (ADTs).

Type Environments
$$\in (\mathbb{X} \xrightarrow{\text{fin}} \mathbb{T}) \times (\mathbb{X}_t \xrightarrow{\text{fin}} (\mathbb{X} \xrightarrow{\text{fin}} \mathbb{T}^*))$$
 (TypeEnv)
$$\Gamma(t) = x_1(\tau_{1,1}, \dots, \tau_{1,m_1}) + \dots + x_n(\tau_{n,1}, \dots, \tau_{n,m_n})$$

and sum types are commutative:

$$\Gamma(\mathtt{A}) = \mathtt{B}(\mathtt{bool}) + \mathtt{C}(\mathtt{num})$$
 equivalent to $\Gamma(\mathtt{A}) = \mathtt{C}(\mathtt{num}) + \mathtt{B}(\mathtt{bool})$

```
case class TypeEnv(
  vars: Map[String, Type] = Map(),
  tys: Map[String, Map[String, List[Type]]] = Map()
) {
  def addVar(pair: (String, Type)): TypeEnv = TypeEnv(vars + pair, tys)
  def addVars(pairs: Iterable[(String, Type)]): TypeEnv =
        TypeEnv(vars ++ pairs, tys)
  def addType(tname: String, ws: Map[String, List[Type]]): TypeEnv =
        TypeEnv(vars, tys + (tname -> ws))
}
```



For example, consider the following an ADT for binary trees:

```
/* ATFAE */
enum Tree {
  case Leaf(Number)
  case Node(Tree, Number, Tree)
} ...
```





For example, consider the following an ADT for binary trees:

```
/* ATFAE */
enum Tree {
  case Leaf(Number)
  case Node(Tree, Number, Tree)
} ...
```

We can add the type information of the Tree ADT to an existing type environment Γ (or tenv) as follows:

```
\Gamma[\mathtt{Tree} = \mathtt{Leaf}(\mathtt{num}) + \mathtt{Node}(\mathtt{Tree}, \mathtt{num}, \mathtt{Tree})]
```

```
val newTEnv = tenv.addType(NameT("Tree"), Map(
   "Leaf" -> List(NumT),
   "Node" -> List(NameT("Tree"), NumT, NameT("Tree"))
))
```



```
/* ATFAE */
enum Tree {
  case Leaf(Number)
  case Node(Tree, Number, Tree)
}
def f(t: Tree): Tree = t
...
```

It is a well-typed ATFAE expression.





```
/* ATFAE */
enum Tree {
  case Leaf(Number)
  case Node(Tree, Number, Tree)
}
def f(t: Tree): Tree = t
...
```

It is a well-typed ATFAE expression.

```
/* ATFAE */
def f(t: Tree): Tree = t
...
```

How about this?



```
/* ATFAE */
enum Tree {
  case Leaf(Number)
  case Node(Tree, Number, Tree)
}
def f(t: Tree): Tree = t
...
```

It is a well-typed ATFAE expression.

```
/* ATFAE */
def f(t: Tree): Tree = t
...
```

How about this? No!

It is **syntactically correct** but the Tree type is **not defined**.



```
/* ATFAE */
enum Tree {
  case Leaf(Number)
  case Node(Tree, Number, Tree)
}
def f(t: Tree): Tree = t
...
```

It is a well-typed ATFAE expression.

```
/* ATFAE */
def f(t: Tree): Tree = t
...
```

How about this? No!

It is **syntactically correct** but the Tree type is **not defined**.

We need to check the **well-formedness** of types with **type environment**.



We need to check the **well-formedness** of types with **type environment**:

```
def mustValid(ty: Type, tenv: TypeEnv): Type = ty match
  case NumT => NumT
  case BoolT => BoolT
  case ArrowT(ptys, rty) =>
    ArrowT(ptys.map(mustValid(_, tenv)), mustValid(rty, tenv))
  case NameT(tn) =>
    if (!tenv.tys.contains(tn)) error(s"invalid type name: $tn")
    NameT(tn)
```

Function Definition



```
def typeCheck(expr: Expr, tenv: TypeEnv): Type = expr match
    ...
    case Fun(params, body) =>
        val ptys = params.map(_.ty)
        for (pty <- ptys) mustValid(pty, tenv)
        val rty = typeCheck(body, tenv.addVars(params.map(p => p.name -> p.
        ty)))
        ArrowT(ptys, rty)
```

$$\Gamma \vdash e : \tau$$

$$\tau-\operatorname{Fun}\frac{\Gamma\vdash\tau_{1}\quad\ldots\quad\Gamma\vdash\tau_{n}}{\Gamma\vdash\lambda(x_{1}:\tau_{1},\ldots,x_{n}:\tau_{n})\cdot e:(\tau_{1},\ldots,\tau_{n})\to\tau}$$

We need to check the well-formedness of parameter types.

Recursive Function Definition



```
def typeCheck(expr: Expr, tenv: TypeEnv): Type = expr match
...
case Rec(name, params, rty, body, scope) =>
  val ptys = params.map(_.ty)
  for (pty <- ptys) mustValid(pty, tenv)
  mustValid(rty, tenv)
  val fty = ArrowT(ptys, rty)
  val bty = typeCheck(body, tenv.addVar(name -> fty)
      .addVars(params.map(p => p.name -> p.ty)))
  mustSame(bty, rty)
  typeCheck(scope, tenv.addVar(name -> fty))
```

$$\frac{\lceil \vdash e : \tau \rceil}{\Gamma \vdash \tau_{1} \quad \dots \quad \Gamma \vdash \tau_{n} \quad \Gamma \vdash \tau}$$

$$\tau - \text{Rec} \quad \frac{\Gamma[x_{0} : (\tau_{1}, \dots, \tau_{n}) \to \tau, x_{1} : \tau_{1}, \dots, x_{n} : \tau_{n}] \vdash e : \tau}{\Gamma[x_{0} : (\tau_{1}, \dots, \tau_{n}) \to \tau] \vdash e' : \tau'}$$

$$\frac{\Gamma[x_{0} : (\tau_{1}, \dots, \tau_{n}) \to \tau] \vdash e' : \tau'}{\Gamma \vdash \text{def } x_{0}(x_{1} : \tau_{1}, \dots, x_{n} : \tau_{n}) : \tau = e; \ e' : \tau'}$$

We need to check the **well-formedness** of parameter and return types.

Function Application



```
def typeCheck(expr: Expr, tenv: TypeEnv): Type = expr match
...
case App(fun, args) => typeCheck(fun, tenv) match
    case ArrowT(ptys, retTy) =>
        if (ptys.length != args.length) error("arity mismatch")
        (ptys zip args).map((p, a) => mustSame(typeCheck(a, tenv), p))
        retTy
    case ty => error(s"not a function type: ${ty.str}")
```

$$au-{ t App} \ rac{\Gamma dash e_0 : (au_1, \dots, au_n) o au}{\Gamma dash e_0 (e_1, \dots, e_n) : au} \ \dots \ \Gamma dash e_n : au_n}{\Gamma dash e_0 (e_1, \dots, e_n) : au}$$

No change in the type checking for function application.

Algebraic Data Types



```
def typeCheck(expr: Expr, tenv: TypeEnv): Type = expr match
...
    case TypeDef(tn, ws, body) =>
        val newTEnv = tenv.addType(tn, ws.map(w => w.name -> w.ptys).toMap)
    for (w <- ws; pty <- w.ptys) mustValid(pty, newTEnv)
    typeCheck(
        body,
        newTEnv.addVars(ws.map(w => w.name -> ArrowT(w.ptys, NameT(tn))))
    )
```

$$\tau - \text{TypeDef} \xrightarrow{\Gamma' = \Gamma[t = x_1(\tau_{1,1}, \dots, \tau_{1,m_1}) + \dots + x_n(\tau_{n,1}, \dots, \tau_{n,m_n})]}{\Gamma' \vdash \tau_{1,1} \quad \dots \quad \Gamma' \vdash \tau_{n,m_n} \quad \Gamma' \begin{bmatrix} x_1 : (\tau_{1,1}, \dots, \tau_{1,m_1}) \to t, \\ \dots, \\ x_n : (\tau_{n,1}, \dots, \tau_{n,m_n}) \to t \end{bmatrix} \vdash e : \tau}$$

$$\tau - \text{TypeDef} \xrightarrow{\Gamma' \vdash \text{enum } t} \begin{cases} \text{case } x_1(\tau_{1,1}, \dots, \tau_{1,m_1}) \\ \dots, \\ \text{case } x_n(\tau_{n,1}, \dots, \tau_{n,m_n}) \end{cases} ; e : \tau}$$

Algebraic Data Types



```
def typeCheck(expr: Expr, tenv: TypeEnv): Type = expr match
...
    case TypeDef(tn, ws, body) =>
        val newTEnv = tenv.addType(tn, ws.map(w => w.name -> w.ptys).toMap)
        for (w <- ws; pty <- w.ptys) mustValid(pty, newTEnv)
        typeCheck(
            body,
            newTEnv.addVars(ws.map(w => w.name -> ArrowT(w.ptys, NameT(tn))))
        )
```

$$\tau - \texttt{TypeDef} \xrightarrow{ \begin{array}{c} \Gamma' = \Gamma[t = x_1(\tau_{1,1}, \ldots, \tau_{1,m_1}) + \ldots + x_n(\tau_{n,1}, \ldots, \tau_{n,m_n})] \\ \Gamma' \vdash \tau_{1,1} & \ldots & \Gamma' \vdash \tau_{n,m_n} & \Gamma' \begin{bmatrix} x_1 : (\tau_{1,1}, \ldots, \tau_{1,m_1}) \to t, \\ \ldots, \\ x_n : (\tau_{n,1}, \ldots, \tau_{n,m_n}) \to t \end{bmatrix} \vdash e : \tau \\ \hline \Gamma \vdash \texttt{enum} \ t \ \begin{cases} \texttt{case} \ x_1(\tau_{1,1}, \ldots, \tau_{1,m_1}) \\ \ldots \\ \texttt{case} \ x_n(\tau_{n,1}, \ldots, \tau_{n,m_n}) \end{cases} \end{cases} ; \ e : \tau \\ \hline \end{array} \right\} ; \ e : \tau$$

It is indeed **type unsound**, and we will fix it later in this lecture.

Pattern Matching



```
def typeCheck(expr: Expr, tenv: TypeEnv): Type = expr match
...
case Match(expr, cs) => typeCheck(expr, tenv) match
  case NameT(tn) =>
    val ts = tenv.tys.getOrElse(tn, error(s"unknown type: $tn"))
    val xs = cs.map(_.name).toSet
    if (ts.keySet != xs || xs.size != cs.length) error("invalid case")
    cs.map { case MatchCase(x, ps, b) =>
        typeCheck(b, tenv.addVars(ps zip ts(x)))
    }.reduce((lty, rty) => { mustSame(lty, rty); lty })
    case _ => error("not a variant")
```

$$\tau-\mathtt{Match} \begin{array}{c} \Gamma \vdash e: t \qquad \Gamma(t) = x_1(\tau_{1,1}, \dots, \tau_{1,m_1}) + \dots + x_n(\tau_{n,1}, \dots, \tau_{n,m_n}) \\ \tau-\mathtt{Match} \end{array} \\ \frac{\forall 1 \leq i \leq n. \; \Gamma_i = \Gamma[x_{i,1}: \tau_{i,1}, \dots, x_{i,m_i}: \tau_{i,m_i}] \qquad \Gamma_1 \vdash e_1: \tau \qquad \dots \qquad \Gamma_n \vdash e_n: \tau}{\Gamma \vdash e \; \mathtt{match} \; \left\{ \begin{array}{c} \mathtt{case} \; x_1(x_{1,1}, \dots, x_{1,m_1}) \Rightarrow e_1 \\ \dots \\ \mathtt{case} \; x_n(x_{n,1}, \dots, x_{n,m_n}) \Rightarrow e_n \end{array} \right\} : \tau}{\Gamma \vdash e \; \mathtt{match}} \end{array}$$

Contents



Type Checker and Typing Rules
 Type Environment for ADTs
 Well-Formedness of Types
 (Recursive) Function Definition and Application
 Algebraic Data Types
 Pattern Matching

2. Type Soundness of ATFAE

Recall: Type Soundness
Algebraic Data Types - Revised (1)

Algebraic Data Types - Revised (2)

Recall: Type Soundness



Definition (Type Soundness)

A type system is sound if it guarantees that a well-typed program will never cause a type error at run-time.





A type system is sound if it guarantees that a well-typed program will never cause a type error at run-time.

Consider the following ATFAE expression:





A type system is sound if it guarantees that a well-typed program will never cause a type error at run-time.

Consider the following ATFAE expression:

It throws a **type error** when evaluating true + 1 at run-time while this expression is **well-typed** (i.e., **unsound type system**).





A type system is sound if it guarantees that a well-typed program will never cause a type error at run-time.

Consider the following ATFAE expression:

It throws a **type error** when evaluating true + 1 at run-time while this expression is **well-typed** (i.e., **unsound type system**).

It happens because the **same type name** A is defined twice and **shadows** the previous one with **different types** for its **variants**.





A type system is sound if it guarantees that a well-typed program will never cause a type error at run-time.

Consider the following ATFAE expression:

It throws a **type error** when evaluating true + 1 at run-time while this expression is **well-typed** (i.e., **unsound type system**).

It happens because the **same type name** A is defined twice and **shadows** the previous one with **different types** for its **variants**.

Let's **forbid** the redefinition of **same type name** in the scope of **ADTs**!

Algebraic Data Types - Revised (1)



```
def typeCheck(expr: Expr, tenv: TypeEnv): Type = expr match
    ...
    case TypeDef(tn, ws, body) =>
        if (tenv.tys.contains(tn)) error(s"already defined type: $tn")
    val newTEnv = tenv.addType(tn, ws.map(w => w.name -> w.ptys).toMap)
    for (w <- ws; pty <- w.ptys) mustValid(pty, newTEnv)
    typeCheck(
        body,
        newTEnv.addVars(ws.map(w => w.name -> ArrowT(w.ptys, NameT(tn))))
    )
```

$$\Gamma' = \Gamma[t = x_1(\tau_{1,1}, \dots, \tau_{1,m_1}) + \dots + x_n(\tau_{n,1}, \dots, \tau_{n,m_n})]$$

$$t \notin \mathsf{Domain}(\Gamma) \qquad \Gamma' \vdash \tau_{1,1} \qquad \Gamma' \vdash \tau_{n,m_n}$$

$$\Gamma' \left[\begin{array}{c} x_1 : (\tau_{1,1}, \dots, \tau_{1,m_1}) \to t, \\ \dots, \\ x_n : (\tau_{n,1}, \dots, \tau_{n,m_n}) \to t \end{array} \right] \vdash e : \tau$$

$$\tau - \mathsf{TypeDef} \qquad \left\{ \begin{array}{c} \mathsf{case} \ x_1(\tau_{1,1}, \dots, \tau_{1,m_1}) \\ \dots, \\ \mathsf{case} \ x_n(\tau_{n,1}, \dots, \tau_{n,m_n}) \end{array} \right\}; \ e : \tau$$









It throws a **type error** when evaluating true + 1 at run-time while this expression is **well-typed** (i.e., **unsound type system**).





It throws a **type error** when evaluating true + 1 at run-time while this expression is **well-typed** (i.e., **unsound type system**).

Since the second A type does not shadow the first one, the type system allows the definition of the second A type.





It throws a **type error** when evaluating true + 1 at run-time while this expression is **well-typed** (i.e., **unsound type system**).

Since the second A type does not shadow the first one, the type system allows the definition of the second A type.

It happens because the first A type **escapes its scope** and is still visible in the scope of the second A type.





It throws a **type error** when evaluating true + 1 at run-time while this expression is **well-typed** (i.e., **unsound type system**).

Since the second A type does not shadow the first one, the type system allows the definition of the second A type.

It happens because the first A type **escapes its scope** and is still visible in the scope of the second A type.

Let's **forbid** the escape of **ADTs** from their scope!

Algebraic Data Types - Revised (2)



```
def typeCheck(expr: Expr, tenv: TypeEnv): Type = expr match
...
    case TypeDef(tn, ws, body) =>
        if (tenv.tys.contains(tn)) error(s"already defined type: $tn")
    val newTEnv = tenv.addType(tn, ws.map(w => w.name -> w.ptys).toMap)
    for (w <- ws; pty <- w.ptys) mustValid(pty, newTEnv)
    mustValid(typeCheck(
        body,
        newTEnv.addVars(ws.map(w => w.name -> ArrowT(w.ptys, NameT(tn))))
    ), tenv)
```

$$\tau^{\prime} = \Gamma[t = x_{1}(\tau_{1,1}, \dots, \tau_{1,m_{1}}) + \dots + x_{n}(\tau_{n,1}, \dots, \tau_{n,m_{n}})]$$

$$t \notin \mathsf{Domain}(\Gamma) \quad \Gamma^{\prime} \vdash \tau_{1,1} \quad \dots \quad \Gamma^{\prime} \vdash \tau_{n,m_{n}}$$

$$\tau^{\prime} \begin{bmatrix} x_{1} : (\tau_{1,1}, \dots, \tau_{1,m_{1}}) \to t, \\ \dots, \\ x_{n} : (\tau_{n,1}, \dots, \tau_{n,m_{n}}) \to t \end{bmatrix} \vdash e : \tau \qquad \Gamma \vdash \tau$$

$$\tau - \mathsf{TypeDef} \quad \frac{\mathsf{Case} \ x_{1}(\tau_{1,1}, \dots, \tau_{1,m_{1}})}{\mathsf{\Gamma} \vdash \mathsf{enum} \ t} \begin{cases} \mathsf{case} \ x_{1}(\tau_{1,1}, \dots, \tau_{n,m_{n}}) \end{cases} ; \ e : \tau$$

$$\mathsf{case} \ x_{n}(\tau_{n,1}, \dots, \tau_{n,m_{n}}) \end{cases} ; \ e : \tau$$

Exercise #12



- Please see this document¹ on GitHub.
 - Implement typeCheck function.
 - Implement interp function.
- It is just an exercise, and you don't need to submit anything.
- However, some exam questions might be related to this exercise.

¹https://github.com/ku-plrg-classroom/docs/tree/main/cose212/atfae.

Summary



1. Type Checker and Typing Rules

Type Environment for ADTs
Well-Formedness of Types
(Recursive) Function Definition and Application
Algebraic Data Types
Pattern Matching

2. Type Soundness of ATFAE

Recall: Type Soundness

Algebraic Data Types - Revised (

Algebraic Data Types - Revised (1)

Algebraic Data Types - Revised (2)

Next Lecture



• Parametric Polymorphism

Jihyeok Park
 jihyeok_park@korea.ac.kr
https://plrg.korea.ac.kr