Lecture 26 – Type Inference (2)

COSE212: Programming Languages

Jihyeok Park



2023 Fall

Recall



- Type inference is the process of automatically inferring the types of expressions.
- We have seen three examples to learn how the type inference works.

```
/* RFAE */ def sum(x) = if (x < 1) 0 else x + sum(x - 1); sum
```

```
/* RFAE */ val app = n => f => f(n); app(42)(x => x)
```

```
/* RFAE */ val id = x => x; val n = id(42); val b = id(true); b
```

- In this lecture, let's learn the details of the type inference algorithm.
- TIFAE TRFAE with type inference.
 - Type Checker and Typing Rules with Type Inference
 - Interpreter and Natural Semantics

Contents



1. Type Checker and Typing Rules with Type Inference

Solutions for Type Constraints

Numbers

Additions

Conditionals

Immutable Variable Definitions and Identifier Lookup

Function Definitions

Recursive Function Definitions

Function Applications

2. Type Unification

Type Resolving

Occurrence Checking

Type Unification

3. Type Inference with Let-Polymorphism

Type Generalization

Type Instantiation

Contents



1. Type Checker and Typing Rules with Type Inference

Solutions for Type Constraints

Numbers

Additions

Conditionals

Immutable Variable Definitions and Identifier Lookup

Function Definitions

Recursive Function Definitions

Function Applications

2. Type Unification

Type Resolving Occurrence Checking Type Unification

3. Type Inference with Let-Polymorphism

Type Generalization

Type Checker and Typing Rules



Let's **1** design **typing rules** of TIFAE 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.

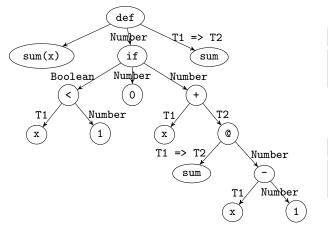
We will keep track of the **variable types** using a **type environment** Γ as a mapping from variable names to their types.

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

Recall: Example 1 – sum



In addition, we need to keep track of the **solution** for **type constraints** over **type variables** to infer the types of expressions.



Type Environment

Type Liviloinnent	
X	\mathbb{T}
х	T1
sum	T1 => T2

Solution

\mathbb{X}_{α}	\mathbb{T}
T1	Number
T2	Number

Solutions for Type Constraints



A **solution** is a mapping from **type variables** to **types**.

Solutions
$$\psi \in \Psi = \mathbb{X}_{\alpha} \xrightarrow{\mathsf{fin}} (\mathbb{T} \uplus \{\bot\})$$
 (Solution)

Type Variables $\alpha \in \mathbb{X}_{\alpha}$

Note that \bot (None) represents a **not yet solved** (free) type variable.

Now, we have new forms of type checker and typing rules.

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

$$\lceil \mathsf{\Gamma}, \psi \vdash e : au, \psi
vert$$

Similar to the memory passing in MFAE for mutation, we will pass the solution ψ and update it during type checking.

(Int)

Numbers



```
def typeCheck(e: Expr, tenv: TypeEnv, sol: Solution): (Type, Solution) = e match
    ...
    case Num(n) => (NumT, sol)
```

$$\Gamma, \psi \vdash e : \tau, \psi$$

$$\tau-\mathtt{Num}\ \overline{\Gamma,\psi \vdash \mathit{n} : \mathtt{num},\psi}$$

Additions



```
def typeCheck(e: Expr, tenv: TypeEnv, sol: Solution): (Type, Solution) = e match
...
case Add(1, r) =>
  val (lty, sol1) = typeCheck(1, tenv, sol)
  val (rty, sol2) = typeCheck(r, tenv, sol1)
  val sol3 = unify(lty, NumT, sol2)
  val sol4 = unify(rty, NumT, sol3)
  (ty, sol4)
```

$$|\Gamma, \psi \vdash e : \tau, \psi|$$

The unify function that takes two types must be the same and updates the solution. We will see how it works later.

Conditionals



```
def typeCheck(e: Expr, tenv: TypeEnv, sol: Solution): (Type, Solution) = e match
...
case If(c, t, e) =>
  val (cty, sol1) = typeCheck(c, tenv, sol)
  val (tty, sol2) = typeCheck(t, tenv, sol1)
  val (ety, sol3) = typeCheck(e, tenv, sol2)
  val sol4 = unify(cty, BoolT, sol3)
  val sol5 = unify(tty, ety, sol4)
  (tty, sol5)
```

$$|\Gamma,\psi \vdash e : au,\psi|$$

$$\tau - \text{If} \ \frac{\Gamma, \psi \vdash e_c : \text{bool}, \psi_c \quad \Gamma, \psi_c \vdash e_t : \tau_t, \psi_t \quad \Gamma, \psi_t \vdash e_e : \tau_e, \psi_e}{\text{unify}(\tau_c, \text{bool}, \psi_e) = \psi' \quad \text{unify}(\tau_t, \tau_e, \psi') = \psi''}{\Gamma, \psi \vdash \text{if} \ (e_c) \ e_t \ \text{else} \ e_e : \tau_t, \psi''}$$





```
def typeCheck(e: Expr, tenv: TypeEnv, sol: Solution): (Type, Solution) = e match
    ...

case Val(x, e, b) =>
    val (ety, sol1) = typeCheck(e, tenv, sol)
    typeCheck(b, tenv + (x -> ety), sol1)

case Id(x) => tenv.getOrElse(x, error(s"free identifier: $x"))
```

$$\Gamma, \psi \vdash e : \tau, \psi$$

$$\tau - \mathtt{Val} \ \frac{\Gamma, \psi_0 \vdash e_1 : \tau_1, \psi_1 \qquad \Gamma[x : \tau_1] \vdash e_2 : \tau_2, \psi_2}{\Gamma, \psi_0 \vdash \mathtt{val} \ x = e_1; \ e_2 : \tau_2, \psi_2}$$

$$\tau$$
-Id $\frac{x \in \mathsf{Domain}(\Gamma)}{\Gamma, \psi \vdash x : \Gamma(x), \psi}$

Function Definitions



```
def typeCheck(e: Expr, tenv: TypeEnv, sol: Solution): (Type, Solution) = e match
...
    case Fun(p, b) =>
        val (pty, sol1) = newTypeVar(sol)
        val (rty, sol2) = typeCheck(b, tenv + (p -> pty), sol1)
        (ArrowT(pty, rty), sol2)
```

$$\Gamma, \psi \vdash e : \tau, \psi$$

$$\tau-\text{Fun }\frac{\alpha_{\textit{p}}\notin\psi\quad \Gamma[\textit{x}:\alpha_{\textit{p}}],\psi[\alpha_{\textit{p}}\mapsto\bot]\vdash\textit{e}:\tau,\psi'}{\Gamma,\psi\vdash\lambda\textit{x}.\textit{e}:\alpha_{\textit{p}}\to\tau,\psi'}$$





```
def typeCheck(e: Expr, tenv: TypeEnv, sol: Solution): (Type, Solution) = e match
...
case Rec(f, p, b, s) =>
  val (pty, sol1) = newTypeVar(sol)
  val (rty, sol2) = newTypeVar(sol1)
  val fty = ArrowT(pty, rty)
  val tenv1 = tenv + (f -> fty)
  val tenv2 = tenv1 + (p -> pty)
  val (bty, sol3) = typeCheck(b, tenv2, sol2)
  val sol4 = unify(bty, rty, sol3)
  typeCheck(s, tenv1, sol4)
```

$$\lceil \mathsf{\Gamma}, \psi \vdash \mathsf{e} : \tau, \psi \rceil$$

$$\begin{aligned} \alpha_{p}, \alpha_{r} \notin \psi & \alpha_{p} \neq \alpha_{r} & \Gamma_{1} = \Gamma[x_{f} \mapsto (\alpha_{p} \to \alpha_{r})] \\ \Gamma_{2} = \Gamma_{1}[x_{p} \mapsto \alpha_{p}] & \Gamma_{2}, \psi[\alpha \mapsto \bot, \alpha' \mapsto \bot] \vdash e_{b} : \tau_{b}, \psi_{b} \\ & \underbrace{\text{unify}(\tau_{b}, \alpha_{r}, \psi_{b}) = \psi_{r}} & \Gamma_{1}, \psi_{r} \vdash e_{s} : \tau_{s}, \psi_{s} \end{aligned}$$

$$\tau - \text{Rec} \frac{ }{ \Gamma_{1}(x_{p} \mapsto \alpha_{p}) + \Gamma_{2}(x_{p}) = e_{b}; e_{s} : \tau_{s}, \psi_{s} }$$

Function Applications



```
def typeCheck(e: Expr, tenv: TypeEnv, sol: Solution): (Type, Solution) = e match
...
    case App(f, a) =>
        val (fty, sol1) = typeCheck(f, tenv, sol)
        val (aty, sol2) = typeCheck(a, tenv, sol1)
        val (rty, sol3) = newTypeVar(sol2)
        val sol4 = unify(ArrowT(aty, rty), fty, sol3)
        (rty, sol4)
```

$$|\Gamma,\psi \vdash e: au,\psi|$$

$$\tau-\texttt{App} \ \frac{ \begin{matrix} \Gamma,\psi \vdash e_f : \tau_f,\psi_f & \Gamma,\psi_f \vdash e_a : \tau_a,\psi_a \\ \hline \alpha_r \notin \psi_a & \texttt{unify}(\tau_a \to \alpha_r,\tau_f,\psi_a[\alpha_r \mapsto \bot]) = \psi' \\ \hline \Gamma,\psi \vdash e_1(e_2) : \alpha_r,\psi' \end{matrix}$$

Contents



1. Type Checker and Typing Rules with Type Inference

Solutions for Type Constraints

Numbers

Addition

Conditionals

Immutable Variable Definitions and Identifier Lookup

Function Definitions

Recursive Function Definitions

Function Applications

2. Type Unification

Type Resolving
Occurrence Checking
Type Unification

Type Inference with Let-Polymorphism

Type Generalization

Type Instantiation

Type Unification



Definition (Type Unification)

Type unification is the process of updating a solution to make two types equal. If the types are not unifiable, then this process fails and throws an exception.

$$\mathtt{unify}: (\mathbb{T} \times \mathbb{T} \times \Psi) \rightharpoonup \Psi$$

For example, if we unify a type variable α and the number type num, the empty solution \varnothing is updated to $[\alpha \mapsto \text{num}]$.

$$\mathtt{unify}(\alpha,\mathtt{num},\varnothing) = [\alpha \mapsto \mathtt{num}]$$

Before, we define the type unification, we need to define the **type resolving** and **occurrence checking** functions.

- **1** Type resolving is the process of recursively resolving a type variable to its representative type to deal with the **type aliasing**.
- Occurrence checking is the process of checking whether a type variable occurs in a type to detect cyclic types.

Type Resolving



To understand why we need the **type resolving** function, let's consider the following example:

Unfortunately, we cannot know that α_2 are num with the solution ψ_3 .

We need to **resolve** the type variable α_1 to find its **representative type** and update its solution to num to deal with the **type aliasing**.

$$\texttt{unify(resolve}(\alpha_1, \psi_2), \texttt{num}, \psi_2) = \psi_3' = \begin{bmatrix} & & & \\ \mathbb{X}_{\alpha} & \mathbb{T} & \\ \hline \alpha_1 & \alpha_2 & \\ \hline \alpha_2 & \texttt{num} & \end{bmatrix}$$

Type Resolving



We can define the **type resolving** function as follows:

$$oxed{\mathsf{resolve}: (\mathbb{T} imes \mathsf{\Psi}) o \mathbb{T}}$$

$$\mathtt{resolve}(\tau,\psi) = \left\{ \begin{array}{ll} \mathtt{resolve}(\tau',\psi) & \text{if } \tau = \alpha \wedge \psi(\alpha) = \tau' \\ \tau & \text{otherwise} \end{array} \right.$$

and implement it in Scala as follows:

```
def resolve(ty: Type, sol: Solution): Type = ty match
  case VarT(k) => sol(k) match
  case Some(ty) => resolve(ty, sol)
  case None => ty
  case _ => ty
```

Occurrence Checking



Let's understand why we need the **occurrence checking** function:

$$\mathtt{unify}(\alpha_1,\mathtt{num} \to \alpha_1,\psi) = \psi'$$

It actually fails because the type variable α_1 occurs in the type $\mathtt{num} \to \alpha_1$, which means it requires **cyclic types** not supported in our type system.

Let's define the **occurrence checking** function as follows:

$$\mathtt{occur}: (\mathbb{X}_{\alpha} \times \mathbb{T} \times \Psi) \to \mathtt{bool}$$

$$\mathtt{occur}(\alpha, \tau, \psi) = \left\{ \begin{array}{ll} \mathtt{true} & \mathsf{if} \ \tau = \alpha \\ \mathtt{occur}(\alpha, \tau_p, \psi) \vee \mathtt{occur}(\alpha, \tau_r, \psi) & \mathsf{if} \ \tau = (\tau_p \to \tau_r) \\ \mathtt{false} & \mathsf{otherwise} \end{array} \right.$$

and implement it in Scala as follows:

```
def occurs(k: Int, ty: Type, sol: Solution): Boolean = resolve(ty, sol) match
  case VarT(l) => k == 1
  case ArrowT(pty, rty) => occurs(k, pty, sol) || occurs(k, rty, sol)
  case _ => false
```

Type Unification



Using the **type resolving** and **occurrence checking** functions, we could define the **type unification** as a partial function:

$$\boxed{\mathtt{unify}: (\mathbb{T} \times \mathbb{T} \times \Psi) \rightharpoonup \Psi}$$

$$\begin{aligned} & \text{unify}(\tau_1,\tau_2,\psi) = \\ & \begin{cases} \psi & \text{if } \tau_1' = \text{num} \wedge \tau_2' = \text{num} \\ \psi & \text{if } \tau_1' = \text{bool} \wedge \tau_2' = \text{bool} \\ \text{unify}(\tau_{1,r},\tau_{2,r},\text{unify}(\tau_{1,p},\tau_{2,p},\psi)) & \text{if } \tau_1' = (\tau_{1,p} \to \tau_{1,r}) \wedge \tau_2' = (\tau_{2,p} \to \tau_{2,r}) \\ \psi & \text{if } \tau_1' = \alpha = \tau_2' \\ \psi[\alpha \mapsto \tau_1'] & \text{if } \tau_1' = \alpha \wedge \neg \text{occur}(\alpha,\tau_2') \\ \psi[\alpha \mapsto \tau_1'] & \text{if } \tau_2' = \alpha \wedge \neg \text{occur}(\alpha,\tau_1') \end{cases}$$

where $\tau_1' = \mathtt{resolve}(\tau_1, \psi)$ and $\tau_2' = \mathtt{resolve}(\tau_2, \psi)$.

- First, it resolves the types τ_1 and τ_2 with the current solution ψ into τ_1' and τ_2' using the **type resolving** function resolve.
- 2 If only one of them $(\tau_1' \text{ or } \tau_2')$ is a type variable, it checks cyclic types using the **occurrence checking** function occur.
- **3** Then, it unifies types τ'_1 and τ'_2 and updates the solution ψ .





And, we can implement the **type unification** function in Scala as follows:

```
def unify(lty: Type, rty: Type, sol: Solution): Solution =
  (resolve(lty, sol), resolve(rty, sol)) match
   case (NumT, NumT) => sol
   case (BoolT, BoolT) => sol
   case (ArrowT(lpty, lrty), ArrowT(rpty, rrty)) =>
     unify(lrty, rrty, unify(lpty, rpty, sol))
   case (VarT(k), VarT(l)) if k == l => sol
   case (VarT(k), rty) if !occurs(k, rty, sol) => sol + (k -> Some(rty))
   case (lty, VarT(k)) if !occurs(k, lty, sol) => sol + (k -> Some(lty))
   case _ => error(s"Cannot unify ${lty.str} and ${rty.str}")
```

```
 \begin{aligned} & \text{unify}(\tau_1,\tau_2,\psi) = \\ & \begin{cases} \psi & \text{if } \tau_1' = \text{num} \wedge \tau_2' = \text{num} \\ \psi & \text{if } \tau_1' = \text{bool} \wedge \tau_2' = \text{bool} \\ & \text{unify}(\tau_{1,r},\tau_{2,r},\text{unify}(\tau_{1,\rho},\tau_{2,\rho},\psi)) & \text{if } \tau_1' = (\tau_{1,\rho} \to \tau_{1,r}) \wedge \tau_2' = (\tau_{2,\rho} \to \tau_{2,r}) \\ \psi & \text{if } \tau_1' = \alpha = \tau_2' \\ \psi[\alpha \mapsto \tau_1'] & \text{if } \tau_1' = \alpha \wedge \neg \text{occur}(\alpha,\tau_2') \\ \psi[\alpha \mapsto \tau_1'] & \text{if } \tau_2' = \alpha \wedge \neg \text{occur}(\alpha,\tau_1') \end{cases}
```

Contents



1. Type Checker and Typing Rules with Type Inference

Solutions for Type Constraints

Numbers

Addition

Conditionals

Immutable Variable Definitions and Identifier Lookup

Function Definitions

Recursive Function Definitions

Function Applications

2. Type Unification

Type Resolving

Occurrence Checking

Type Unification

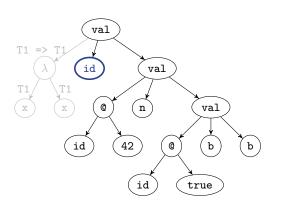
3. Type Inference with Let-Polymorphism

Type Generalization

Type Instantiation

Recall: Example 3 - id





	Type Environment
X	\mathbb{T}
id	[T1] { T1 => T1 }

		301011011
\mathbb{X}_{α}	\mathbb{T}	
T1	_	

C = 1......

Let's **generalize** the type T1 => T1 into a **polymorphic type** for id with **type variable** T1 as a **type parameter**.

We call this **let-polymorphism** because it only introduces polymorphism for the let-binding (e.g., val).

Type Generalization



We can define the **type generalization** function gen as follows:

$$\boxed{ \begin{split} & \text{gen}: (\mathbb{T} \times \mathbb{\Gamma} \times \Psi) \to \mathbb{T}^{\forall} \\ & \text{gen}(\tau, \Gamma, \psi) = \forall \alpha_1, \dots, \alpha_m.\tau \end{split} \quad \text{where} \quad \text{free}_{\tau}(\tau, \psi) \setminus \text{free}_{\Gamma}(\Gamma, \psi) = \{\alpha_1, \dots, \alpha_m\} }$$

and the free type variables in each component as follows:

$$\mathsf{free}_{\tau}: (\mathbb{T} \times \Psi) \to \mathcal{P}(\mathbb{X}_{\alpha}) \\ \mathsf{free}_{\tau}: (\tau, \psi) = \left\{ \begin{array}{ll} \mathsf{free}_{\tau}: (\mathbb{T} \times \Psi) \to \mathcal{P}(\mathbb{X}_{\alpha}) \\ \mathsf{free}_{\tau'}(\tau', \psi) & \mathsf{if} \ \tau = \alpha \wedge \psi(\alpha) = \tau' \\ \mathsf{free}_{\tau_p}(\tau_p, \psi) \cup \mathsf{free}_{\tau_r}(\tau_r, \psi) & \mathsf{if} \ \tau = (\tau_p \to \tau_r) \\ \varnothing & \mathsf{otherwise} \end{array} \right.$$

$$\boxed{ \texttt{free}_{\Gamma} : (\mathbb{\Gamma} \times \Psi) \to \mathcal{P}(\mathbb{X}_{\alpha}) } \\ \texttt{free}_{\Gamma}([x_1 : \tau_1^{\forall}, \dots, x_n : \tau_n^{\forall}], \psi) = \texttt{free}_{\tau_n^{\forall}}(\tau_1^{\forall}, \psi) \cup \dots \cup \texttt{free}_{\tau_n^{\forall}}(\tau_n^{\forall}, \psi) }$$

Immutable Variable Defs. with Type Generalization APLRG

```
def typeCheck(e: Expr, tenv: TypeEnv, sol: Solution): (Type, Solution) = e match
    ...

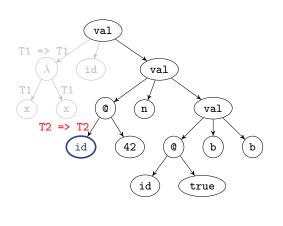
case Val(x, e, b) =>
    val (ety, sol1) = typeCheck(e, tenv, sol)
    val polyty = gen(ety, tenv, sol1)
    typeCheck(b, tenv + (x -> polyty), sol1)
```

$$\Gamma, \psi \vdash e : \tau, \psi$$

$$\tau-\mathrm{Val}\ \frac{\Gamma,\psi_0\vdash e_1:\tau_1,\psi_1\ \ \mathsf{gen}(\tau_1,\Gamma,\psi_1)=\tau_1^\forall\ \ \Gamma[x:\tau_1^\forall]\vdash e_2:\tau_2,\psi_2}{\Gamma,\psi_0\vdash \mathrm{val}\ x=e_1;\ e_2:\tau_2,\psi_2}$$

Recall: Example 3 - id





	Type Environment
\mathbb{X}	T
id	[T1] { T1 => T1 }

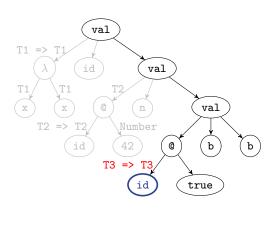
Solution		
\mathbb{X}_{α}	\mathbb{T}	
T1	-	
T2	ı	

Calution

Let's define a new **type variable T2** to **instantiate** the **type variable T1**. And, **substitute T1** with **T2**.

Recall: Example 3 - id





	Type Environment
X	\mathbb{T}
id	[T1] { T1 => T1 }
n	T2

Solution	
\mathbb{X}_{α}	T
T1	_
T2	Number
Т3	-

Let's define a new **type variable T3** to **instantiate** the **type variable T1**. And, **substitute T1** with **T3**.

Type Instantiation



We can define the **type instantiation** function inst as follows:

$$\begin{array}{c} [\texttt{inst}: (\mathbb{T}^\forall \times \Psi) \to (\mathbb{T} \times \Psi)] \\ \\ \texttt{inst}(\forall \alpha_1, \ldots, \alpha_m.\tau, \psi) = (\\ \\ \texttt{subst}(\tau, \psi[\alpha_1 \mapsto \alpha_1', \ldots, \alpha_m \mapsto \alpha_m']), \\ \\ \psi[\alpha_1' \mapsto \bot, \ldots, \alpha_m' \mapsto \bot] \\) \\ \\ \texttt{where} \qquad \alpha_1', \ldots, \alpha_m' \notin \psi \land \forall 1 \leq i < j \leq m. \ \alpha_i' \neq \alpha_j' \end{array}$$

and the **type substitution** function subst as follows:

$$\mathtt{subst}: (\mathbb{T} \times \Psi) \to \mathbb{T}$$

$$\mathtt{subst}(\tau,\psi) = \left\{ \begin{array}{ll} \mathtt{subst}(\tau',\psi) & \text{if } \tau = \alpha \wedge \psi(\alpha) = \tau' \\ \mathtt{subst}(\tau_{\mathit{p}},\psi) \to \mathtt{subst}(\tau_{\mathit{r}},\psi) & \text{if } \tau = (\tau_{\mathit{p}} \to \tau_{\mathit{r}}) \\ \tau & \text{otherwise} \end{array} \right.$$





```
def typeCheck(e: Expr, tenv: TypeEnv, sol: Solution): (Type, Solution) = e match
...

case Id(x) =>
   val ty = tenv.getOrElse(x, error(s"free identifier: $x"))
   inst(ty, sol)
```

$$|\Gamma, \psi \vdash e : \tau, \psi|$$

$$au- ext{Id} \ rac{\Gamma(x) = au^orall \ ext{inst}(au^orall, \psi) = (au, \psi')}{\Gamma, \psi \vdash x : au, \psi'}$$

Summary



1. Type Checker and Typing Rules with Type Inference

Solutions for Type Constraints

Numbers

Additions

Conditionals

Immutable Variable Definitions and Identifier Lookup

Function Definitions

Recursive Function Definitions

Function Applications

2. Type Unification

Type Resolving

Occurrence Checking

Type Unification

3. Type Inference with Let-Polymorphism

Type Generalization

Type Instantiation

Exercise #15



- 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/tifae.

Next Lecture



Course Review

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