Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

Définitions

Théorie des typ

Résultats

Conclusio

Formalisatio

Annexe

# Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

Enogad Le Biavant-Frederic

Alain René Lesage MPI

2025

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

#### Présentation générale

Parsing

I neorie des type

i courtat

\_ ....

Annexe

#### Karm, 2022



Comment optimiser la vitesse de compilation en fusionnant analyse syntaxique et sémantique ?

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse

> syntaxique Enogad Le Biavant–

#### Présentation

générale

Définitions Parsing

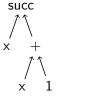
Théorie des typ

Résultat

Conclusio

i Oilliansatii

$$\Gamma = \{ \texttt{+}: int \rightarrow int \rightarrow int \}$$
 let  $succ = \lambda x. (\texttt{+} \ x \ 1)$ 





Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant–

## Présentation

#### générale

Définitions Parsing

Dácultat

Formalisatio

$$\begin{split} \Gamma &= \{ \texttt{+}: int \rightarrow int \rightarrow int, \ \texttt{x}: \tau \} \\ \texttt{let} \ succ &= \lambda x. (\texttt{+} \ x \ 1) \end{split}$$



$$\begin{cases} x:\tau \end{cases} \bigwedge_{\mathsf{X}} \bigwedge_{\mathsf{Y}} + \\ \{x:\tau \} \bigwedge_{\mathsf{Y}} \bigwedge_{\mathsf{Y}} \bigwedge_{\mathsf{Y}} \begin{cases} x:\tau \end{cases}$$

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

#### Présentation générale

Définitions Parsing

Théorie des typ

Résultats

00...0....

Formalisatio

$$\Gamma = \{ \texttt{+} : int \rightarrow int \rightarrow int, \ \texttt{x} : \tau \}$$
 let  $succ = \lambda x. (\texttt{+} \ x \ 1)$ 



Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

#### Présentation générale

Définitions

Parsing Théorie des types

Résultats

Anneve

$$\Gamma = \{ \texttt{+}: int \rightarrow int, \ \texttt{x}: \tau \}$$
 let  $succ = \lambda x. (\texttt{+} \ x \ 1)$ 

Parsing récursif descendant

succ 
$$\{x:\tau\} / \bigwedge \{\tau = int\}$$
 
$$\times + \{x:\tau\} / \bigwedge$$
 
$$\times 1$$

#### Grammaire

```
Optimisation
de la vitesse
de compilation
par la fusion
entre inférence
de types et
analyse
syntaxique
```

Enogad Le Biavant– Frederic

Présentation générale

80.10.010

Parsing

Théorie des types

Resultats

Formalisatio

```
program
                 expr
                 abs
expr
                 app
                 letbinding
            := term [\{ term \}]
app
abs
            := "\" id "." expr
               "let" id "=" expr "in" expr
letbinding
term
                 string
                 int
                 bool
                 id
                 "(" expr ")"
```

#### TT - Définitions

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

Définitions

Théorie des types

Résultat

Conclusi

Formalisation

Annexe

Définition de Type : classification de termes (Church). Théorie de travail : Lambda calcul simplement typé (LCST), polymorphique.

let 
$$id = \lambda x.x : \forall \sigma \to \sigma$$

#### Hindley-Milner

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation

Définitions

Deminitions

Théorie des types

Résultats

Conclusio

Formalisation

Annovo

$$\frac{x:\sigma\in\Gamma}{\Gamma\vdash x:\sigma} \text{ var }$$

$$\frac{\Gamma, x: \tau \vdash e: \tau'}{\Gamma \vdash \lambda x. e: \tau \rightarrow \tau'} \text{ abs }$$

$$\frac{\Gamma \vdash f: \tau \to \tau' \qquad \Gamma \vdash e: \tau}{\Gamma \vdash f \ e: \tau'} \ \mathsf{app}$$

#### Hindley-Milner

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

générale

Théorie des types

Earmalication

Anneve

#### Algorithme W

- Assignation de variables de types aux expressions
- 2 Génération de contraintes
- **3** Substitutions
- **4** Unification
- 5 Instantiation, généralisation

$$\begin{cases} x:\tau \rbrace / \bigwedge \{expr:\beta \rbrace \\ \times \quad \text{expr} \end{cases}$$
 
$$\{\beta = int, \tau = int \} \uparrow \\ \{x:\tau \rbrace / \bigwedge \\ \times \quad 1 \end{cases}$$

#### Résultats

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

Définitions

Parsing

Théorie des type

Résultats

\_ ....

Formalisatio

Annexe

$$\mathcal{W}: \tilde{\Gamma} \times Expr \rightarrow Subst \times Type \\ \mathcal{W}^*: \tilde{\Gamma} \times L \rightarrow Subst \times \Gamma \times Expr \times L$$

Machine: i7 5th gen 3.00Ghz

Fichier de test: 1000 premiers nombres de church

Version non optimisée :  $\approx 26.37s$ 

Version optimisée :  $\approx 3.88s$ 

Fichier de test : 10K application de fonction successeur

Version non optimisée :  $\approx 1.48s$ 

Version optimisée :  $\approx 1.51s$ 

#### Résultats

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation

générale

Définitions

Théorie des typ

Résultats

Conclus

Formalisation

#### Conclusion

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

generale

Théorie des tyn

Résultat

Conclusion

Formalisation

Annexe

#### Bénéfices

Durée inférieure lorsque beaucoup d'unifications

Plus permissif envers les grammaires très imbriquées que les systèmes traditionnels

#### **Formalisation**

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

Définitions

Théorie des ty

Résultat

Conclusion

Formalisation

Annexe

Automates d'arbres :  $\mathcal{A} = (\mathcal{F}, Q, Q_f, \Delta)$ 

lacktriangleright : Alphabet gradué (fonction d'arité ar)

On peut décrire la grammaire comme une NRTG

#### Annexe - Lexemes

```
Optimisation
 de la vitesse
de compilation
 par la fusion
entre inférence
 de types et
   analyse
  syntaxique
```

```
type literal =
    | Str of string
    Int of int
    | Bool of bool
    [@@deriving show]
type t =
    | Let
     Lambda
    I Dot.
    | Assign
     In
    | LParen
    | RParen
    | Id of string
    | Literal of literal
    [@@deriving show]
type program = t list
[@@deriving show]
```

#### Annexe - Expressions

```
Optimisation
de la vitesse
de compilation
par la fusion
entre inférence
de types et
analyse
syntaxique
```

Enogad Le Biavant– Frederic

Présentation générale

Définitions

Théorie des type

\_ . .

Farma dia asi a

```
type literal =
    | Str of string
    | Int of int
    | Bool of bool
[@@deriving show]

type t =
    | Var of string
    | Abs of string * t
    | App of t * t
    | Let of string * t * t
    | Literal of literal
[@@deriving show]
```

```
Optimisation
 de la vitesse
de compilation
 par la fusion
entre inférence
 de types et
    analyse
  syntaxique
```

```
type t =
   (* Primitives *)
    I Bool
    1 Str
    | Int
   | TVar of string
   (* Function *)
   | Abs of t * t (* @-> *)
   (* Scheme *)
    | Forall of string * t
    [@@deriving show];;
let (Q \rightarrow) t1 t2 = Abs (t1, t2)
let counter = ref 0::
let fresh tv () =
   let v = "t" ^ string_of_int !counter in
   incr counter;
   TVar v
let reset_tv_counter () = counter := 0;;
```

```
Optimisation
 de la vitesse
de compilation
 par la fusion
entre inférence
 de types et
   analyse
 syntaxique
```

```
module TypeMap = Map.Make(String)
type subst = t TypeMap.t (* // *)
let rec apply_subst (s: subst) (t: t) =
   match t with
    | TVar v -> begin
       match TypeMap.find_opt v s with (* If type is not found,
            works like Id *)
        I Some x \rightarrow x
        | None -> t
   end
    | Abs (arg, ret) -> Abs (apply_subst s arg, apply_subst s
        ret)
    | Forall (v, t) -> Forall (v, apply_subst (TypeMap.remove v
        s) t)
    l t. -> t.
let ( *\&* ) s1 s2 =
   TypeMap.union (fun _ t _ -> Some t) (TypeMap.map
         (apply_subst s1) s2) s1
```

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

Définitions Parsing

Théorie des type

Résultat

Formalisatio

```
type env = t TypeMap.t;; (* String : Type *)
let extend_env env k t = TypeMap.add k t env
let apply_env (env: env) k: t =
    try TypeMap.find k env
    with Not_found -> failwith ("Var not in scope : " ^ k)
let apply_subst_to_env subst env = TypeMap.map (fun t ->
        apply_subst subst t) env;;
```

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

Définitions

Théorie des type

Résultats

Conclusio

Formalisation

```
module Ftv = Set.Make (String);;

type ftvs = Ftv.t

let rec free_tvs t: ftvs =
    match t with
    | Int | Bool | Str -> Ftv.empty
    | TVar v -> Ftv.singleton v
    | Abs (t1, t2) -> Ftv.union (free_tvs t1) (free_tvs t2)
    | Forall (v, t) -> Ftv.remove v (free_tvs t)

let free_tvs_env env =
    TypeMap.fold (fun _ t acc -> Ftv.union (free_tvs t) acc) env
    Ftv.empty
```

```
Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

Enogad Le Biavant-Frederic
```

Présentation générale

Definitions Parsing

Theorie des type

Formalisation

```
exception TypeError of string
let rec unify t1 t2: subst =
   match t1, t2 with
   | Bool, Bool | Int, Int | Str, Str -> TypeMap.empty
   | TVar v1, TVar v2 when v1 = v2 -> TypeMap.empty
   | TVar v, t | t, TVar v ->
       if Ftv.mem v (free tvs t) then
          raise (TypeError ("Occurs check failed for variable "
               ^ v))
       else
          TypeMap.singleton v t
   | Abs (t1a, t1b), Abs (t2a, t2b) ->
       let s1 = unify t1a t2a in
       let s2 = unify (apply_subst s1 t1b) (apply_subst s1 t2b)
           in
       s2 *&* s1
   | Forall (v1, t1), Forall (v2, t2) ->
       let t2' = apply_subst (TypeMap.singleton v2 (TVar v1)) t2
           in
       unify t1 t2'
   | _ -> raise (TypeError "Cannot unify types")
                                                               18 / 36
```

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

Définitions

Théorie des types

Resultats

Formalication

```
let generalize env t =
   let env fv = free tvs env env in
   let ftvs_of_t = free_tvs t in
   let vars_to_generalize = Ftv.filter (fun v -> not (Ftv.mem v
        env fv)) ftvs of t in
   match Ftv.elements vars_to_generalize with
   | [] -> t.
   | vars -> List.fold_right (fun v acc -> Forall (v, acc))
        vars t
let rec instantiate t =
   match t with
   | Forall (v, t) -> let fresh var = fresh tv () in
        instantiate (apply_subst (TypeMap.singleton v fresh_var)
        t)
   l -> t
```

```
Optimisation
de la vitesse
de compilation
par la fusion
entre inférence
de types et
analyse
syntaxique
```

Enogad Le Biavant– Frederic

Présentation générale

Définitions Parsing

Dácultato

\_ . .

Formalisation

```
type traversal = {
   expr: Expr.t;
   rest: Token.program;
let match literals = function
   | Token.Str s -> Expr.Literal (Expr.Str s)
   | Token.Int i -> Expr.Literal (Expr.Int i)
   | Token.Bool b -> Expr.Literal(Expr.Bool b)
let rec parse_expr (program: program): traversal =
   match program with
   | Lambda :: Id arg :: Dot :: rest ->
       let { expr = body: rest = rest': } = parse expr rest in
           expr = Abs (arg, body);
           rest = rest':
   | Let :: _ -> parse_let program
     _ -> parse_app program
```

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

D/finitions

Parsing Théorie des types

Résultats

Conclusio

Formalisation

```
and parse_let (program: program): traversal =
  match program with
  | Let :: Id i :: Assign :: t ->
     let { expr = body; rest; } = parse_expr t in

  let { expr = in_expr; rest = rest'; } = parse_ins rest in
  {
     expr = Let (i, body, in_expr);
     rest = rest';
  }
  | _ -> failwith "Expected 'Let'."
```

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

Définitions

Théorie des types

Résultats

Formalisation

```
and parse_ins (program: program): traversal =
   match program with
   | In :: t -> parse_expr t
   | _ -> failwith "No 'in' clause given. "

and parse_parenthesized (program: program): traversal =
   match program with
   | LParen :: t -> begin
    let e = parse_expr t in
    match e.rest with
   | RParen :: t' -> { expr = e.expr; rest = t'; }
   | _ -> failwith "Unclosed parenthesis"
   end
   | _ -> failwith "Expected LParen."
```

match rest with

| -> begin

and parse\_app (program: program): traversal =

let { expr = lhs\_expr; rest; } = lhs in

expr = App (lhs\_expr, rhs\_expr);

let { expr = rhs\_expr; rest = rest'; } = parse\_term

| [] | In :: \_ | RParen :: \_ -> 1hs

let rec aux (lhs: traversal) =

rest in

```
Optimisation
 de la vitesse
de compilation
 par la fusion
entre inférence
 de types et
   analyse
 syntaxique
```

```
rest = rest';
                            } in
                            aux res
                         end
                     in
                     let initial = parse_term program in
Annexe
                     aux initial
```

let res = {

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

Définitions

Parsing Théorie des types

Résultats

\_ .....

i ormansatioi

```
and parse_term (program: program): traversal =
   match program with
   | Id x :: rest ->
          expr = Var x;
          rest;
   | Token Literal | :: t ->
          expr = match_literals 1;
          rest = t;
   | LParen :: _ -> parse_parenthesized program
   | RParen :: _ -> failwith "Alone RParen. ?"
   | _ -> failwith "Unexpected token"
```

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

D.(C., M.

Théorie des type

Résultate

F. ......

```
let parse (program: program) =
        let { expr; rest = _; } = parse_expr program in
    expr
;;

let type_of_literal = function
    | Expr.Str _ -> Types.Str
    | Expr.Bool _ -> Types.Bool
    | Expr.Int _ -> Types.Int
;;
```

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

Définitions Parsing

Dámiltoto

Formalisatio

```
let rec infer_types_aux env e cont =
   let open Types in
   match e with
   | Expr.Literal 1 ->
       let t = type_of_literal l in
       cont (t, TypeMap.empty)
   | Var x ->
       let t = apply_env env x in
       cont (t, TypeMap.empty)
   | Abs (x, e) ->
       let t = fresh_tv () in
       let env' = extend env env x t in
       infer_types_aux env' e (fun (t', s) ->
           let t_res = (apply_subst s t) @-> t' in
           cont (t res. s)
```

```
Optimisation
de la vitesse
de compilation
par la fusion
entre inférence
de types et
analyse
syntaxique
```

Enogad Le Biavant– Frederic

Présentation générale

Définitions Parsing

Récultate

. . .

Formalisatio

```
| App (e1, e2) ->
   infer_types_aux env e1 (fun (t1, s1) ->
       let env' = apply_subst_to_env s1 env in
       infer_types_aux env' e2 (fun (t2, s2) ->
          let tv = fresh_tv () in
          let s3 = unify (apply_subst s2 t1) (t2 @-> tv) in
          let f subst = s3 *&* s2 *&* s1 in
          cont (apply_subst f_subst tv, f_subst)
Let (x, e1, e2) ->
   infer_types_aux env e1 (fun (t1, s1) ->
       let t' = generalize env (apply_subst s1 t1) in
       let env' = extend_env env x t' in
       infer_types_aux env' e2 (fun (t2, s2) ->
          cont (t2, s2 *&* s1)
       )
```

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

D.(C., M.

Théorie des type

Dácultat

E......

```
let infer_types env e =
        infer_types_aux env e (fun res -> res)

let run (p: program) (env: Types.env) =
    let ast = parse p in
    let t, _ = infer_types env ast in
    (ast, t)

;;
```

type traversal = {

t: Types.t;

(tv. env')

expr: Expr.t;

subst: Types.subst;

rest: Token.t list:

let match\_literals = function

| Token.Str s -> Literal (Str s) | Token.Int i -> Literal (Int i) | Token.Bool b -> Literal(Bool b)

let tv = Types.fresh\_tv () in

let env' = Types.extend\_env env id tv in

let fresh\_id (id: string) (env: Types.env): Types.t \* Types.env =

4□ > 4□ > 4 亘 > 4 亘 > □ ● 9 Q (P)

29/36

```
Optimisation
 de la vitesse
de compilation
 par la fusion
entre inférence
 de types et
   analyse
  syntaxique
Annexe
```

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

generale

Théorie des types

Résultats

Conclusion

Formalisatio

```
let rec parse_expr (program: Token.t list) (env: Types.env):
    traversal =
   match program with
    | Lambda :: Id arg :: Dot :: rest ->
       let (tv. env') = fresh id arg env in
       let { expr = body; t; subst; rest = rest'; } = parse_expr
            rest env' in
       {
           expr = Abs (arg, body);
           t = Types.Abs (Types.apply_subst subst tv, t);
           subst:
           rest = rest':
    | Let :: _ -> parse_let program env
    | _ -> parse_app program env
```

```
Optimisation
de la vitesse
de compilation
par la fusion
entre inférence
de types et
analyse
syntaxique
```

Biavant– Frederic

Présentation générale

Définitions

Parsing

Résultats

Formalisatio

```
and parse_let (program: Token.t list) (env: Types.env):
    traversal =
   match program with
   | Let :: Id i :: Assign :: t ->
       let open Types in
       let { expr = body; t = t1; subst = subst1; rest; } =
           parse_expr t env in
       let t' = generalize env (apply_subst subst1 t1) in (* Get
           the type of the let def *)
       let env' = extend_env env i t' in (* Add the typesig of
           the let-expr to the ctx *)
       let { expr = in_expr; t = t2; subst = subst2; rest =
           rest'; } = parse_ins rest env' in
          expr = Let (i, body, in_expr);
          t = t2:
          subst = subst2 *&* subst1;
          rest = rest':
   | _ -> failwith "Expected 'Let'."
                                                                31/36
```

```
Optimisation
de la vitesse
de compilation
par la fusion
entre inférence
de types et
analyse
syntaxique
```

Biavant– Frederic

Présentation générale

Définitions

Théorie des type

Résultats

Formalisatio

```
and parse_ins (program: Token.t list) (env: Types.env):
    traversal =
   match program with
   | In :: t -> parse_expr t env
   | _ -> failwith "No 'in' clause given. "
and parse_parenthesized (program: Token.t list) (env:
    Types.env): traversal =
   match program with
   | LParen :: t -> begin
       let e = parse_expr t env in
       match e.rest with
       | RParen :: t' -> { expr = e.expr; t = e.t; subst =
           e.subst: rest = t': }
       _ -> failwith "Unclosed parenthesis"
   end
   | _ -> failwith "Expected LParen."
```

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

Définitions Parsing

Récultate

Formalisatio

```
and parse_app (program: Token.t list) (env: Types.env):
    traversal =
   let open Types in
   let rec aux (lhs: traversal) =
       let { expr = lhs_expr; t = t1; rest; subst = s1; } = lhs
           in
       match rest with
         [] | In :: | RParen :: -> lhs
       | _ -> begin
          let { expr = rhs_expr; t = t2; subst = s2; rest =
               rest'; } = parse_term rest env in
          let tv = fresh_tv () in
          let s3 = unify (apply_subst s2 t1) (Abs (t2, tv)) in
          let f_subst = s3 *&* s2 *&* s1 in
```

```
Optimisation
de la vitesse
de compilation
par la fusion
entre inférence
de types et
analyse
syntaxique
```

Enogad Le Biavant– Frederic

Présentation générale

générale

Parsing Théorie des types

Résultats

Farma dianata

```
let res = {
        expr = App (lhs_expr, rhs_expr);
        t = apply_subst f_subst tv;
        subst = f_subst;
        rest = rest';
        } in
        aux res
    end
in
let initial = parse_term program env in
aux initial
```

```
Optimisation
 de la vitesse
de compilation
 par la fusion
entre inférence
 de types et
   analyse
  syntaxique
Annexe
```

```
and parse_term (program: Token.t list) (env: Types.env):
    traversal =
   match program with
   | Id x :: rest. ->
       let t = Types.apply_env env x in
          expr = Var x:
          t;
          subst = Types.TypeMap.empty;
          rest:
   | Token Literal | :: t ->
          expr = match_literals 1;
          t = type_literals 1;
          subst = Types.TypeMap.empty;
          rest = t;
   | LParen :: _ -> parse_parenthesized program env
   | RParen :: _ -> failwith "Alone RParen. ?"
   -> failwith "Unexpected token"
                                       4 D > 4 A > 4 B > 4 B > B
```

Optimisation de la vitesse de compilation par la fusion entre inférence de types et analyse syntaxique

> Enogad Le Biavant– Frederic

Présentation générale

générale

Parsing

Résultats

Conclusion

Formalisatio

```
let parse (program: Token.program) (env: Types.env) =
   Types.reset_tv_counter ();
   let { expr; t; subst = _; rest = _; } = parse_expr program
        env in
   (expr, t)
```