Coccinelle for Rust
https://gitlab.inria.fr/coccinelle/coccinelleforrust.git

Julia Lawall, Tathagata Roy

July 23, 2023

Goals

- Perform repetitive transformations at a large scale.

Goals

- Perform repetitive transformations at a large scale.
  - Rust is 1.6 MLOC.
  - The Linux kernel is 23 MLOC.

Goals

- Perform repetitive transformations at a large scale.
  - Rust is 1.6 MLOC.
  - The Linux kernel is 23 MLOC.
  - Collateral evolutions: a change in an API requires changes in all clients.

Goals

- Perform repetitive transformations at a large scale.
  - Rust is 1.6 MLOC.
  - The Linux kernel is 23 MLOC.
  - Collateral evolutions: a change in an API requires changes in all clients.
- Provide a transformation language that builds on developer expertise.

Goals

- Perform repetitive transformations at a large scale.
  - Rust is 1.6 MLOC.
  - The Linux kernel is 23 MLOC.
  - Collateral evolutions: a change in an API requires changes in all clients.
- Provide a transformation language that builds on developer expertise.
- Changes + developer familiarity = (semantic) patches

An example change (Rust repository)

```
commit d822b97a27e50f5a091d2918f6ff0ffd2d2827f5
Author: Kyle Matsuda <kyle.yoshio.matsuda@gmail.com>
Date:   Mon Feb 6 17:48:12 2023 -0700

    change usages of type_of to bound_type_of

diff --git a/compiler/rustc_borrowck/src/diagnostics/conflict_errors.rs b/compiler/.../conflict_errors.rs
@@ -2592,4 +2592,4 @@ fn annotate_argument_and_return_for_borrow(
             } else {
-                let ty = self.infcx.tcx.type_of(self.mir_def_id());
+                let ty = self.infcx.tcx.bound_type_of(self.mir_def_id()).subst_identity();
                 match ty.kind() {
                     ty::FnDef(_, _) | ty::FnPtr(_) => self.annotate_fn_sig(
diff --git a/compiler/rustc_borrowck/src/diagnostics/mod.rs b/compiler/.../mod.rs
@@ -1185,4 +1185,4 @@ fn explain_captures(
                         matches!(tcx.def_kind(parent_did), rustc_hir::def::DefKind::Impl { .. })
                             .then_some(parent_did)
-                            .and_then(|did| match tcx.type_of(did).kind() {
+                            .and_then(|did| match tcx.bound_type_of(did).subst_identity().kind() {
                                 ty::Adt(def, ..) => Some(def.did()),
...
```

136 files changed, 385 insertions(+), 262 deletions(-)

An example change (Rust repository)

```
commit d822b97a27e50f5a091d2918f6ff0ffd2d2827f5
Author: Kyle Matsuda <kyle.yoshio.matsuda@gmail.com>
Date:   Mon Feb 6 17:48:12 2023 -0700

    change usages of type_of to bound_type_of

diff --git a/compiler/rustc_borrowck/src/diagnostics/conflict_errors.rs b/compiler/.../conflict_errors.rs
@@ -2592,4 +2592,4 @@ fn annotate_argument_and_return_for_borrow(
             } else {
-                let ty = self.infcx.tcx.type_of(self.mir_def_id());
+                let ty = self.infcx.tcx.bound_type_of(self.mir_def_id()).subst_identity();
                 match ty.kind() {
                     ty::FnDef(_, _) | ty::FnPtr(_) => self.annotate_fn_sig(
diff --git a/compiler/rustc_borrowck/src/diagnostics/mod.rs b/compiler/.../mod.rs
@@ -1185,4 +1185,4 @@ fn explain_captures(
                         matches!(tcx.def_kind(parent_did), rustc_hir::def::DefKind::Impl { .. })
                             .then_some(parent_did)
-                            .and_then(|did| match tcx.type_of(did).kind() {
+                            .and_then(|did| match tcx.bound_type_of(did).subst_identity().kind() {
                                 ty::Adt(def, ..) => Some(def.did()),
...
```

136 files changed, 385 insertions(+), 262 deletions(-)

Creating a semantic patch: Step 1: remove irrelevant code

```
-                 let ty = self.infcx.tcx.type_of(self.mir_def_id())
+                 let ty = self.infcx.tcx.bound_type_of(self.mir_def_id()).subst_identity()
```

```
-                     .and_then(|did| match tcx.type_of(did).kind() {
+                     .and_then(|did| match tcx.bound_type_of(did).subst_identity().kind() {
```

Creating a semantic patch: Step 2: abstract over subterms using metavariables

```
@@

@@

- self.infcx.tcx.type_of(self.mir_def_id())
+ self.infcx.tcx.bound_type_of(self.mir_def_id()).subst_identity()
```

Creating a semantic patch: Step 2: abstract over subterms using metavariables

```
@@
expression tcx, arg;
@@

- tcx.type_of(arg)
+ tcx.bound_type_of(arg).subst_identity()
```

Creating a semantic patch: Step 2: abstract over subterms using metavariables

```
@@
expression tcx, arg;
@@

- tcx.type_of(arg)
+ tcx.bound_type_of(arg).subst_identity()
```

Updates over 200 call sites.

An outlier

```
            let (shim_size, shim_align, _kind) = ecx.get_alloc_info(alloc_id);
+           let def_ty = ecx.tcx.bound_type_of(def_id).subst_identity();
            let extern_decl_layout =
-               ecx.tcx.layout_of(ty::ParamEnv::empty().and(ecx.tcx.type_of(def_id))).unwrap();
+               ecx.tcx.layout_of(ty::ParamEnv::empty().and(def_ty)).unwrap();
            if extern_decl_layout.size != shim_size || extern_decl_layout.align.abi != shim_align {
                throw_unsup_format!(
                    "`extern` static `{name}` from crate `{krate}` has been declared \
```

An outlier

```
            let (shim_size, shim_align, _kind) = ecx.get_alloc_info(alloc_id);
+           let def_ty = ecx.tcx.bound_type_of(def_id).subst_identity();
            let extern_decl_layout =
-               ecx.tcx.layout_of(ty::ParamEnv::empty().and(ecx.tcx.type_of(def_id))).unwrap();
+               ecx.tcx.layout_of(ty::ParamEnv::empty().and(def_ty)).unwrap();
            if extern_decl_layout.size != shim_size || extern_decl_layout.align.abi != shim_align {
                throw_unsup_format!(
                    "`extern` static `{name}` from crate `{krate}` has been declared \
```

The developer has created a new name to avoid a long line.

- Could address it manually.
- Could create a rule for the special case of nested function call contexts (probably not worth it for one case).

Alternate semantic patch

```
@@
expression tcx, arg;
@@

  tcx.
-     type_of(arg)
+     bound_type_of(arg).subst_identity()
```

Putting tcx in the context ensures any comments will be preserved.

An example: change in context

```
commit 1ce80e210d152619caa99b1bc030f57a352b657a
Author: Oliver Scherer <oli-obk@users.noreply.github.com>
Date:   Thu Feb 16 09:25:11 2023 +0000

    Allow `LocalDefId` as the argument to `def_path_str`

diff --git a/compiler/rustc_borrowck/src/lib.rs b/compiler/rustc_borrowck/src/lib.rs
@@ -124,3 +124,3 @@ pub fn provide(providers: &mut Providers) {
 fn mir_borrowck(tcx: TyCtxt<'_>, def: LocalDefId) -> &BorrowCheckResult<'_> {
     let (input_body, promoted) = tcx.mir_promoted(def);
-    debug!("run query mir_borrowck: {}", tcx.def_path_str(def.to_def_id()));
+    debug!("run query mir_borrowck: {}", tcx.def_path_str(def));
diff --git a/compiler/rustc_hir_analysis/src/check/check.rs b/compiler/rustc_hir_analysis/src/check/check.rs
@@ -494,5 +494,5 @@ fn check_item_type(tcx: TyCtxt<'_>, id: hir::ItemId) {
     debug!(
         "check_item_type(it.def_id={:?}, it.name={})",
         id.owner_id,
-        tcx.def_path_str(id.owner_id.to_def_id())
+        tcx.def_path_str(id.owner_id)
     );
...
```

18 files changed, 68 insertions(+), 54 deletions(-)

An example: change in context

Want to drop `.to_def_id()` but only in an argument to `tcx.def_path_str`:

```
@@
expression tcx, arg;
@@

-       tcx.def_path_str(arg.to_def_id())
+       tcx.def_path_str(arg)
```

Updates 48 call sites in 18 files.

An example: multiple cases

```
commit 298ae8c721102c36243335653e57a7f94e08f94a
Author: Michael Goulet <michael@errs.io>
Date:   Wed Feb 22 22:23:10 2023 +0000

    Rename ty_error_with_guaranteed to ty_error, ty_error to ty_error_misc

diff --git a/compiler/rustc_borrowck/src/region_infer/opaque_types.rs b/compiler/.../opaque_types.rs
@@ -156,3 +156,3 @@ pub(crate) fn infer_opaque_types(
                     });
-                    prev.ty = infcx.tcx.ty_error_with_guaranteed(guar);
+                    prev.ty = infcx.tcx.ty_error(guar);
                 }
@@ -248,3 +248,3 @@ fn infer_opaque_definition_from_instantiation(
         if let Some(e) = self.tainted_by_errors() {
-            return self.tcx.ty_error_with_guaranteed(e);
+            return self.tcx.ty_error(e);
         }
...
diff --git a/compiler/rustc_hir_analysis/src/astconv/mod.rs b/compiler/rustc_hir_analysis/src/astconv/mod.rs
@@ -429,2 +429,2 @@ fn provided_kind(
                         self.inferred_params.push(ty.span);
-                        tcx.ty_error().into()
+                        tcx.ty_error_misc().into()
```

32 files changed, 121 insertions(+), 140 deletions(-)

An example: multiple cases

Two changes:

- From `ty_error_with_guaranteed` to `ty_error` (1 argument)
- From `ty_error` to `ty_error_misc` (no arguments)

```
@@
expression tcx, arg;
@@
- tcx.ty_error_with_guaranteed(arg)
+ tcx.ty_error(arg)

@@
expression tcx, arg;
@@
- tcx.ty_error()
+ tcx.ty_error_misc()
```

An example: searching for variants

```
commit f3f9d6dfd92dfaeb14df891ad27b2531809dd734
Author: Eduard-Mihai Burtescu <edy.burt@gmail.com>
Date:   Fri Jun 14 00:48:52 2019 +0300

    Unify all uses of 'gcx and 'tcx.

diff --git a/src/librustc/infer/error_reporting/mod.rs b/src/librustc/infer/error_reporting/mod.rs
@@ -460,6 +460,6 @@ impl<'gcx, 'tcx> Printer<'gcx, 'tcx> for AbsolutePathPrinter<'gcx, 'tcx> {
             type DynExistential = !;
             type Const = !;

-            fn tcx<'a>(&'a self) -> TyCtxt<'gcx, 'tcx> {
+            fn tcx<'a>(&'a self) -> TyCtxt<'tcx> {
                 self.tcx
             }
@@ -1977,4 +1976,4 @@ pub fn enter_global<'gcx, F, R>(gcx: &'gcx GlobalCtxt<'gcx>, f: F) -> R
     pub unsafe fn with_global<F, R>(f: F) -> R
     where
-        F: for<'gcx, 'tcx> FnOnce(TyCtxt<'gcx, 'tcx>) -> R,
+        F: for<'tcx> FnOnce(TyCtxt<'tcx>) -> R,
     {
```

341 files changed, 3109 insertions(+), 3327 deletions(-)

An example: searching for variants

A first attempt:

```
@rule type@
@@
- TyCtxt<'gcx, 'tcx>
+ TyCtxt<'tcx>
```

An example: searching for variants

A first attempt:

```
@rule type@
@@
- TyCtxt<'gcx, 'tcx>
+ TyCtxt<'tcx>
```

This does part of the work, but some change sites are overlooked:

- `TyCtxt<'tcx, 'tcx>`, `TyCtxt<'_, '_>`
- `DepNodeParams<'gcx, 'tcx>`

An example: searching for variants

A first attempt:

```
@rule type@
@@
- TyCtxt<'gcx, 'tcx>
+ TyCtxt<'tcx>
```

This does part of the work, but some change sites are overlooked:

- `TyCtxt<'tcx, 'tcx>`, `TyCtxt<'_, '_>`
- `DepNodeParams<'gcx, 'tcx>`
- And others?

An example: searching for variants

A more general attempt:

```
@rule type@
identifier Ty;
@@
- Ty<'gcx, 'tcx>
+ Ty<'tcx>
```

An example: searching for variants

A more general attempt:

```
@rule type@
identifier Ty;
@@
- Ty<'gcx, 'tcx>
+ Ty<'tcx>
```

How to find other change sites, like `TyCtxt<'tcx, 'tcx>`, `TyCtxt<'_, '_>`:

- Want to change all uses of types that are somewhere used with `'gcx`.

An example: searching for variants

A more general attempt:

```
@r type@
identifier Ty;
@@
- Ty<'gcx, 'tcx>
+ Ty<'tcx>

@rule type@
identifier r.Ty;
@@
(
- Ty<'tcx, 'tcx>
+ Ty<'tcx>
|
- Ty<'_, '_>
+ Ty<'_>
)
```

Summary: Features seen so far

- Semantic patches:
  Patch-like transformation specification, abstracted using metavariables.
- Multiple rules/rule ordering.
- Inheritance.
- Disjuctions.

All of these features are implemented!

Some more Coccinelle features

- `...`
  - `A ... B`: `A` can reach `B` through the function control-flow.
  - `<... A ...>`, `<+... A ...+>`: `A` occurs 0+ or 1+ times.

Some more Coccinelle features

- `...`
  - `A ... B`: `A` can reach `B` through the function control-flow.
  - `<... A ...>`, `<+... A ...+>`: `A` occurs 0+ or 1+ times.
- Position variables.
- Script code.

Some more Coccinelle features

- `...`
  - `A ... B`: `A` can reach `B` through the function control-flow.
  - `<... A ...>`, `<+... A ...+>`: `A` occurs 0+ or 1+ times.
- Position variables.
- Script code.
- Constraints on metavariables (types, etc.).
- Fresh identifiers.
- $*$ for matching without transformation.

These features are coming soon...

Some Coccinelle internals

Input: Parsing provided by Rust Analyzer.

- Used both for Rust code and for semantic patch code.
- Will provide type inference, when needed.

Some Coccinelle internals

Input: Parsing provided by Rust Analyzer.

- Used both for Rust code and for semantic patch code.
- Will provide type inference, when needed.

Output: Pretty printing provided by `rustfmt`.

- Preserves comments.

Some Coccinelle internals

In the middle:

- Wrap Rust code and semantic patch code, eg to indicate metavariables.
- Match semantic patch code against Rust code, to collect change sites and metavariable bindings.
- On a successful match, apply the changes, instantiated according to the metavariable bindings, reparse, and repeat with the next rule.

Practical issues

```
Usage: main [OPTIONS] --coccifile <COCCIFILE> --targetpath <TARGETPATH>

Options:
  -c, --coccifile <COCCIFILE>            Path of Semantic Patch File path
  -t, --targetpath <TARGETPATH>          Path of Rust Target file/folder path
  -o, --output <OUTPUT>                  Path of transformed file path
  -r, --rustfmt-config <RUSTFMT_CONFIG>  rustfmt config file path [default: rustfmt.toml]
  -i, --ignore <IGNORE>                  [default: ]
  -d, --debug-cocci
      --apply
      --suppress-diff
      --suppress-formatting
      --no-parallel
  -h, --help                             Print help
  -V, --version                          Print version
```

Conclusion

- Transformation on atomic terms completed (expressions, types, etc).
- Transformation on terms connected by a control-flow path (...) in progress.
- Small-scale testing has been done:
  - Replicating real changes on real Rust code.
- Patchparse extended to Rust, to find test cases at a larger scale.
- Availability: https://gitlab.inria.fr/coccinelle/coccinelleforrust.git