

Experience Report: Rapid Reengineering of Legacy Software using Java Reflection

Blair Durkee and Daniel T. Welch and Murali Sitaraman

Technical Report RSRG-17-02

School of Computing

100 McAdams

Clemson University

Clemson, SC 29634-0974 USA

January 2017

Copyright © 2017 by the authors. All rights reserved.

Experience Report: Rapid Reengineering of Legacy Software using Java Reflection

Blair Durkee, Daniel T. Welch, Murali Sitaraman

School of Computing, Clemson University, South Carolina, SC, 29634, USA

Abstract. The objective of this paper is to describe our experience in rapidly reengineering legacy software through the application of Java reflection in a research compiler. This process would be of value to language evolution and development efforts, in general, and reengineering efforts, in particular. It was performed in the course of a software project that has slowly evolved from a simple language translator to a sophisticated verifying compiler with multiple back-ends over a 20-year period, even as the language itself has continued to change with research progress. To minimize the temporal costs of refactoring legacy code in the face of continuous evolution of the language and the need for new additional components (e.g., translators for multiple languages) for various uses in our project, we used a reflection-based, generic walker based on a proper implementation of the visitor pattern. This paper describes the dynamic walker and our experience in employing the walker for code generation.

Keywords: compiler; design pattern; language translation; reengineering; reflection; visitor

1 Introduction

Reusability and maintainability are key nonfunctional characteristics of well-engineered software. These qualities which are especially important for long-running research software projects are often sacrificed in the process of rapidly prototyping solutions to cutting edge research questions. Typically reengineering is necessary to improve the existing software to continue its successful evolution. This paper discusses a reengineering solution to a language development project that aims to make it easier for continued evolution. The solution also serves as another practical example of the role of reflection in software projects with significant legacy code.

The compiler and associated tools in the reengineering effort presented in this paper have been developed over the course of many years and successive generations of developers. These tools have evolved and changed significantly, yet many early design decisions still persist. One example is the mechanism used to traverse the abstract syntax tree (AST) that represents parsed code. As with any compiler, this logic is a key component used in multiple stages of compilation such as pre-processing, population, analyzing, semantic checking, translation, etc. The initial implementation of AST traversal worked sufficiently

well, but it turned out to be an impediment as the initial compiler evolved into a verifying compiler with multiple backend translators. This is the motivation for the reflective tree walker described in this paper. While the idea of a reflection-based visitor is not novel[1][2][3], our experience demonstrates how reflection can be leveraged for efficient reengineering of legacy code. The challenge was to overcome our system’s non-reusable and difficult-to-maintain preexisting tree walking strategy with a solution that meets the standards of reusability and maintainability but does not compromise rate of development.

To give some background and context about this component, we will briefly describe the research context—the RESOLVE compiler. RESOLVE is an integrated programming and specification language that seeks to realize the grand challenge of software verification [4]. Given an implementation for a specification and specifications of reused components, the verifying compiler generates necessary and sufficient verification conditions (VCs) for implementation correctness. The compiler is modular and generates VCs for one component at a time. These VCs are then sent to an integrated prover which attempts to automatically establish the validity of each VC generated using available mathematical results. VCs must be established to ensure program correctness. While this process cannot guarantee that software is specified correctly, it rules out implementation errors by proving programs correct with respect to a given specification. In addition to multiple VC generators, the compiler also includes translators to C and Java. Through a web IDE, the compiler under discussion has been used by researchers [5] and by educators for teaching reasoning principles in a variety of CS courses at Clemson and elsewhere for over 5 years [6].

2 Related Work

Since the emergence of the first compiled, high level languages, ASTs and tree traversal patterns have garnered both an extensive amount of study, and a large (still growing) body of literature [7]. While the concept of ASTs and their usage is well established, AST construction, reusability, and long term maintenance remain relevant topics in the software engineering community. In this section, we consider several existing parsing tools capable of generating ASTs—emphasizing the maintainability cost of using such tools, and relevant mechanisms they might provide for AST traversal.

The GNU “compiler compiler,” Bison¹ [8]—commonly referred to as a “parser generator”—is a tool that produces a state machine capable of recognizing sentences in a given language. By annotating rules of Bison grammars with semantic actions, users are able to construct their own specific AST representations of the underlying grammar. There are several disadvantages with this approach. The first is that Bison does not provide any built in support for automatic AST node construction or tree traversal: Users must create their own nodes, and find their own means of traversal (either with a visitor, or traditional recursive

¹ Bison serves as a modern successor to the original tool, YACC (Yet Another Compiler Compiler) developed in the early 1970s.

‘evaluate’ methods). The second disadvantage is that developers must learn a specific, specialized Bison syntax (separate from its BNF-style grammar syntax) to specify semantic actions. A third disadvantage concerns modularity: The need to annotate grammar rules with specific semantic actions inherently couples a grammar more closely to a given application. This coupling not only decreases the reusability (and generality) of a grammar but also makes it harder to maintain over time, since every change to the rules necessitates further changes to the actions.

Another tool that improves upon the model set by Bison is SableCC [9]. SableCC provides the tools necessary to convert a BNF grammar into Java packages for lexing, parsing, and tree analysis. One particular improvement over Bison is the tool’s ability to produce default Java classes for each node in a user’s AST hierarchy. The analysis package included in the tool defines an abstract class which allows for efficient and foolproof implementation of tree traversal logic. The process of parsing the code and walking the tree is all done by generated code, leaving the developer free to implement visitor methods needed in the analysis and code generation portions of the compiler. However, in order to make use of these features, users are required to annotate the grammar with a separate set of tree transformation rules needed to produce a valid AST. Thus, like Bison, we feel that this ties a grammar too closely to a given intermediate representation.

The final representative tool we discuss is Antlr (Another Tool for Language Recognition). Historically, Antlr has allowed its grammars to be annotated with Bison-style actions and SableCC-style “rewrite operators,” enabling the parser to generate automatically an AST hierarchy, which users could then choose to (manually) outfit with visitor support. However, with more recent iterations of the tool, automatic abstract syntax construction has been removed altogether, in favor of concrete syntax (parse trees). Accompanying this radical change in emphasis is also the addition of a Sax-Dom style event processing mechanism—manifesting itself in the form of automatically generated parse tree listeners and/or visitors².

In terms of the ongoing evolution of the RESOLVE compiler, the implications of these additions are immense³: Not only does Antlr now automatically construct a tree, but it also provides mechanisms for walking and visiting this tree. The allure of not needing to construct your own tree or implement your own traversal patterns is complemented by the tool’s ability to keep grammars completely separate from application specific code (transformations, rewrites, etc). This is made possible by relegating event handling and semantic actions to application-bound parse tree listener and visitor classes.

Despite the maintainability, modularity, and reusability benefits to be had, there remain several difficult (ongoing) decisions regarding this tool’s incorporation into our existing toolchain.

² These features are present in all versions of the tool since Antlr v4

³ The compiler previously used Antlr v3.5

- *Representation concerns*: Antlr, and its automatic walking and visiting mechanisms only operate on Antlr-built parse trees, not custom ASTs. Thus, any switch in representation this drastic at this point in the project would no doubt entail an enormous re-engineering effort across all aspects of our existing system.
- *Performance concerns*: Though reflection-based visitors (as this paper demonstrates) offer improved flexibility and maintenance both in terms of source code (no `accept` methods are required in AST nodes) and overall object structure (seamless integration of new AST nodes into the tree), there is still a concern over performance [7]. These benefits however are offset by the cost of using Java reflection: which is slower than traditional visitors, as our algorithm relies on examining the structure and class hierarchy of our AST nodes at runtime.

Our current system employs a *hybrid* approach: We use Antlr’s new, automatically generated parse tree listening interfaces to construct our AST independent of grammar level tree transformations and actions (upholding reuse and maintainability). We then make use of the tree walking technique described in this paper to traverse our AST. This approach affords us the flexibility to avoid system-wide rewrites, while still taking advantage of the newer features offered by the most recent iteration of the Antlr parsing tool.

3 Initial Implementation and Motivation

The pipeline for compiling high-level code begins with lexing, parsing, and building an abstract syntax tree. This data structure is a logical representation of the source code, and it will be used in nearly every subsequent stage of the compilation pipeline. The need to traverse the tree in each stage presents a problem of code reuse. The traversal mechanism will remain the same during each use, but it must invoke different node visitation logic in each compilation stage (e.g., populator, analyzer, code generator). In order to avoid code duplication, the traversal must be decoupled from the visitation logic, which is a non-trivial task. This task, however, is simplified by a widely accepted design pattern found in the Gang of Four’s seminal work *Design Patterns* [10]. This “visitor pattern” was used in the initial development of our research compiler about 20 years ago, but it was used imperfectly.

Figure 1 provides a snippet of the earlier code to illustrate the problem. The left-hand column shows the “accepting” interfaces and code and right-hand column shows the traversal necessary to visit the statements within the body of a procedure (i.e., method). Note that this figure and the following discussion is not specific to our particular source language, and the same kind of code can be imagined for translating any source language. The design of the tree traversal component’s first iteration bears close resemblance to the visitor pattern: Each class representing a type of node in the syntax tree contains an `accept` method which can dispatch the appropriate logic through subtype polymor-

phism. However, the algorithm for traversal was not properly separated from the data structure as the visitor pattern requires.

The very purpose of the visitor pattern is to decouple a data structure from the logic operating on it. While manifestations of this pattern may vary, they must necessarily adhere to that specific design principle of separation. During the initial development, when design patterns had just been broadly introduced, this principle of separation was not followed cleanly.

```

public abstract
class ResolveConceptualVisitor {
    visitProcedureDecl(
        ProcedureDecl e) {}
}

public abstract
class ResolveConceptualElement {
    abstract void accept(
        ResolveConceptualVisitor e);
}

public class ProcedureDecl
extends ResolveConceptualElement {
    List<Stmt> myStatements;

    public void accept(
        ResolveConceptualVisitor e) {
        v.visitProcedureDecl(this);
    }
}

public class Analyzer
extends ResolveConceptualVisitor {

    public void visitProcedureDecl(
        ProcedureDecl e) {
        table.beginProcedureScope();
        visitStmtList(e.getStatements());
        table.endProcedureScope();
    }

    private void visitStmtList(
        List<Stmt> e) {
        for (Stmt s : e.getStatements()) {
            visitStmt(s);
        }
    }

    public void visitStmt(Stmt e) {
        e.accept(this);
    }
}

```

Fig. 1. An implementation of the visitor pattern that compromises separation of concerns.

The traversal logic, according to visitor pattern, should be contained within the `accept` visitor method. In this initial implementation, it has been moved to the `visit` method contained in the `ResolveConceptualVisitor` component. Consequently, every `ResolveConceptualVisitor` will bear the responsibility of traversing the syntax tree. Thus any change in the tree structure—no matter how minor—will necessitate large, cross-component refactors.

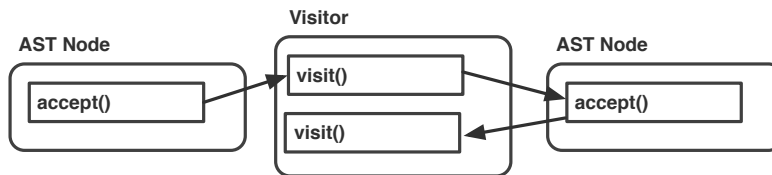


Fig. 2. A high level look at the organization of the flawed visitor.

The reason for this flaw is related to having a singular visit method rather than pre and post visits. Having only one visit method per tree node restricts the traversal to only a pre-ordering or a post-ordering. Figure 1 demonstrates a common scenario for AST traversal—the opening and closing of new scopes in the

symbol table—which requires both orderings. The scope is opened as we traverse down the tree and closed as we traverse back up, but of course, this is not possible if we visit the node only once. The legacy code attempts to tackle this problem by placing the traversal logic inside the visitor methods and, in the process, misapplies the visitor pattern. It is possible to reclaim a proper visitor pattern implementation by adding preorder and postorder visits. Consider the code with that minor change, shown in figure 3. The `ResolveConceptualVisitor` now has fewer methods and the traversal logic is properly decoupled from the visitation logic.

```

public abstract
  class ResolveConceptualVisitor {
    preProcedureDecl(ProcedureDecl e) {}
    postProcedureDecl(ProcedureDecl e) {}
  }

public abstract
  class ResolveConceptualElement {
    abstract void accept(
      ResolveConceptualVisitor e);
  }

public class ProcedureDecl
  extends ResolveConceptualElement {
  List<Stmt> myStatements;

  public void accept(
    ResolveConceptualVisitor e) {
    v.preProcedureDecl(this);
    for (Stmt s : e.getStatements()) {
      s.accept();
    }
    v.postProcedureDecl(this);
  }
}

public abstract class Stmt
  extends ResolveConceptualVisitor {
  abstract void accept(
    ResolveConceptualVisitor e) {}
}

public class Analyzer
  extends ResolveConceptualElement {

  public void preProcedureDecl(
    ProcedureDecl e) {
    table.beginProcedureScope();
  }

  public void postProcedureDecl(
    ProcedureDecl e) {
    table.endProcedureScope();
  }
}

```

Fig. 3. A proper implementation of the visitor pattern.

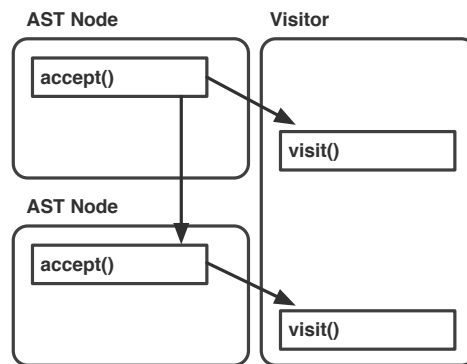


Fig. 4. A high level look at the organization of the fixed visitor.

4 A Dynamic Tree Walker

While the corrections demonstrated in figure 4 could be made, we needed a solution that would not require refactoring of our legacy code. Our existing components needed to continue to work in their present form while we developed new versions of these components. Additionally we believed we could develop a solution that would consume much less time, both initially and in the long-term, than a significant refactoring would. Therefore we decided to pursue a third, even more robust implementation that would exist independently of—and work simultaneously with—legacy code.

The solution involves the creation of two new classes: `TreeWalker` and `TreeWalkerVisitor`. `TreeWalker` is the new traversal mechanism which is completely decoupled from both the tree structure and the visitation logic. It dynamically analyzes the tree composition at runtime using Java reflection. This `TreeWalkerVisitor` replaces the old `ResolveConceptualVisitor` class and provides a new abstract visitor class to implement visitor logic for the various components in the compiler. This design does not modify any existing code and still utilizes the `ResolveConceptualElement` AST classes. This allows the legacy code to continue to work alongside the new dynamic traversal component. In other words, old components can continue to work until new components are ready to be dropped in place.

With this new design, `TreeWalker` will dynamically analyze the structure of the tree at runtime and invoke the appropriate visitor methods as it traverses the tree. Figure 5 is a highly simplified version of the traversal algorithm code (see [Appendix](#) for extended code sample).

```
public void visit(ResolveConceptualElement e) {
    invokeVisitorMethods("pre", e);

    for (ResolveConceptualElement node : e.getChildren()) {
        visit(node);
    }
    invokeVisitorMethods("post", e);
}
```

Fig. 5. A revised `visit` method.

The `visit` method is a simple recursive, depth-first traversal of the AST. At each level, the procedure will make “pre” call before visiting children and a “post” call after. These calls are made via `invokeVisitorMethods`—a local method used to construct the appropriate visitor method name and dispatch the correct calls using reflection techniques. These calls include pre and post methods for each class (base and derived classes) represented by the AST node object.

The dynamic logic for retrieving a node’s children is contained in the `getChildren` defined in the root of the AST class hierarchy, using the base `ResolveConceptualElement` class. This design choice allows for more control over the order of traversal. The default base method uses Java reflection to obtain a list of the children for a

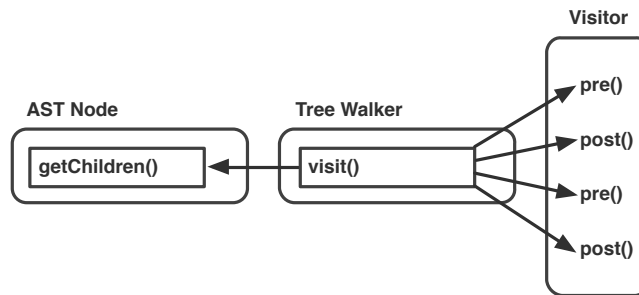


Fig. 6. A revised visit method.

given node and returns the children in a list of unspecified order. Figure 7 shows a simplified version of this dynamic `getChildren` method (see Appendix for extended code sample). If the order is important (or needs to be different from the default), then derived classes can override the method with static logic for returning the children.

```

public List<ResolveConceptualElement> getChildren() {
    List<ResolveConceptualElement> children = new LinkedList<>();
    ArrayList<Field> fields = this.getDeclaredFields();

    for (Field field : fields) {
        if (ResolveConceptualElement.class.isAssignableFrom(field.getType())) {
            children.add(ResolveConceptualElement.class.cast(field.get(this)));
        }
    }

    return children;
}

```

Fig. 7. An implementation of `getChildren`.

Use of the dynamic tree walker is simple and straightforward. To traverse an AST and apply appropriate visitation logic, first create an instance of the `TreeWalker` class, passing an instance of a `TreeWalkerVisitor` (such as the populator or code generator) as a parameter to the constructor. Then, simply call `visit` on the root of the abstract syntax tree.

5 Application Example: Code Generation

One important application of the walking mechanism concerns code generation. Since development began on the RESOLVE compiler, code generation—like the many other phases of compilation—has been constrained by many factors, including the following:

1. *Correct by construction*: Provided with successfully verified RESOLVE source-code, it is the translator’s responsibility to model, as faithfully as possible,

each construct of the source language *within* the target language. This modeling process—performed to maintain the established correctness of the original source—typically precludes the possibility of any sort of syntax-directed translation, as any code generated fitting such a model inevitably ends up looking wildly different from the original source.

2. *Extensibility*: The design of the translator must allow users to relatively easily tweak the output of a given construct, add support for altogether new constructs (accounting for the rapidly developing nature of the source language), and not preclude the addition of any future target languages. This is ultimately one of the reasons we choose to perform source-to-source translation, as opposed generating byte code such as JVM or LLVM single static assignment form directly: It allows a certain level of flexibility—leaving us free to temporarily sidestep the non-trivial problem of developing (and maintaining) a fully blown byte level interpretation of every construct, in favor of more fruitful, verification-related avenues of research.
3. *Reusability*: If two or more supported target languages share similar constructs, the (separate) modules responsible for generating code for each should not duplicate code. Rather, they would ideally be designed to share as much common translation logic as possible, typically via an abstract class or some other means. However, if this is to occur, the translator in question must be designed in such a way that it enforces a strict separation between the logic governing the collection of translation related information, and the actual formatted output of this information.

In this section, by way of a small example, we detail our approach that combines the pre-post visit methods of the tree walker, with the Antrl-authored templating tool, *Stringtemplate*³ [11].

5.1 Stringtemplate Overview

In order to better understand the example that follows, we provide a brief overview of the Stringtemplate language, emphasizing its notation, and usage. A popular ‘definition’ of a template describes it simply as a “document with holes” that users can choose to fill with *attributes*. For instance, the following is a template describing a (RESOLVE) variable declaration.

```
VarDecl(name, type) ::= "Var <name> : <type>";
```

In this case, `name` and `type` serve as attributes to this template, and are to be substituted where they appear within the angled braces (<.>) in the text. In order to interact with this template, users can obtain a reference to it, assign its attributes, and print it out—demonstrated in Listing 1.1.

³ [www.theantlr.org.atlassian.net/wiki/display/ST4/StringTemplate+4+Documentation](http://www.theantlr.org/atlassian.net/wiki/display/ST4/StringTemplate+4+Documentation)

Listing 1.1. Basic template manipulation.

```
//A 'group' is collection of templates defined in an external file
ST variableDecl = myGroup.getInstanceOf("VarDecl");

//We fill in the attributes for 'name' and 'type'
variableDecl.add("name", "X");
variableDecl.add("type", "Integer");

//Then render the overall template (ST) and print the resulting string
System.out.println(variableDecl.render());

//Result:
>$ Var X : Integer;
```

If however we wanted to modify our template to correctly account for declarations involving two or more variables of the same type, we could simply tweak the template's definition to read as follows.

```
VarDecl(name, type) ::= "Var <name> separator = ', ' > : <type>";
```

This small change instructs Stringtemplate to interpret any *multi-valued* attribute passed as a comma delimited list. For instance, if the addition of `name` in the previous example was replaced with,

```
variableDecl.add("name", "X").add("name", "Y");
```

Stringtemplate would yield the following output.

```
>$ Var X, Y : Integer;
```

This section only provides readers with the bare minimum knowledge required to understand the discussion of translation that follows. Readers interested in learning more about this tool are encouraged to refer to [11] or [12].

5.2 RESOLVE Module Translation

To better understand the tree walker's key role in code generation, we consider the trivial operation in Listing 1.2.

Listing 1.2. An example piece of source code.

```
Operation Foo(evaluates J: Integer);
Procedure
    Bar(J);
end Foo;
```

There is little needed to translate `Foo` into Java (or C) for execution. In fact, the translator class hierarchy (depicted in Figure 8) only needs to maintain global reference to a stack of partially filled in, intermediate templates representing various levels traversed of the AST, and the file from which to obtain new, target-language specific templates.

The first step we take in this process is to define a set of templates which, when rendered, outputs the target language representation of the original source (in this example, Java). Listing 1.3 summarizes all the templates we require, including those for functions, parameters, calls, variable arguments, and, at the most granular, (potentially) qualified types.⁴

⁴ It turns out that many templates as they appear in Listing 1.3 (such as `FunctionDefinition`), are suitable for both Java and C output. Thus, to avoid duplication at the template-level, we actually declare templates such as these in a separate template group file that plays the role of an 'abstract-template-group,' from which our dedicated C and Java groups inherit.

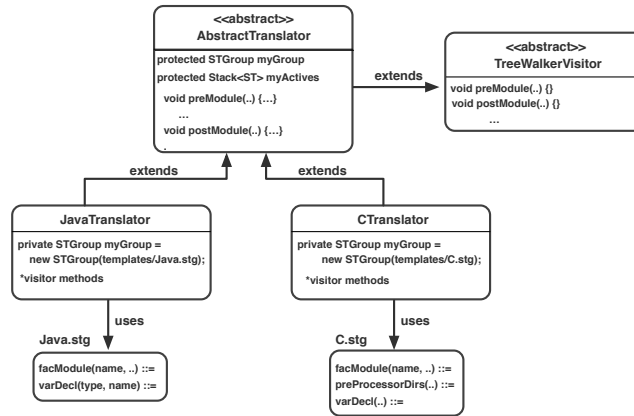


Fig. 8. A translation hierarchy. Each concrete subclass of the abstract translator has an accompanying, dedicated Stringtemplate group file containing all templates tailored to the target language of the class.

Listing 1.3. Templates required for translation of our example. Note: Templates wrapped by “<<. .>>” indicate that whitespace, newlines, and tabs are to be preserved.

```
//functions
FunctionDefTemplate(name, type, params, statements) ::= <<
public <type> <name> (<parameters; sep=', '> {
  <statements; separator='\n'>
}>>

//parameter declarations
ParameterDeclTemplate(type, name) ::= "<type> <name>"

//call statements
CallStmntTemplate(qualifier, name, args) ::= "<if(qualifier)>.<endif><name><args; sep=', '>";

//variable expressions
VarNameExpTemplate(qualifier, name) ::= "<name>"

//named types
NameTypeTemplate(qualifier, name) ::= "<if(qualifier)>.<endif><name>"
```

It’s now up to the Java Translator to completely “fill” the `name`, `return type`, `params`, and `statements` attributes of `FunctionDefTemplate`—either using strings, or other filled-in, templates. One effective way to do so—that we now demonstrate—involves simply using a stack, and the tree walker’s pre-post traversal.

The table depicted in Figure 9 illustrates both the order and state of the translation template stack over the course of the AST’s traversal. At every stage of translation, the top of this stack maintains a reference to the template actively being built, or added to, while the bottom always refers to the result (or, the outermost enclosing template that is to be rendered upon completion).

Thus, translation of a given construct of any target language is performed as follows:

1. In the context of a `preC` visitor method (where `C` corresponds to some source construct), we first retrieve `C`’s corresponding template, `T`, from the current translation subclass’s template group file. We then pass any easily obtainable

Visitor method:	Action:	Stack after action:	Current top-of-stack render progress:
preOperationDecl	<i>push</i>	FunctionDefTemplate	public void Foo(<params; sep=', '>){ <statements; sep= '\n '> }
preParameterDecl	<i>push</i>	ParamDeclTemplate FunctionDefTemplate	<type> J
preNamedType	<i>push</i>	NameTypeTemplate ParamDeclTemplate FunctionDefTemplate	RType
postNamedType	<i>pop</i>	ParamDeclTemplate FunctionDefTemplate	(RType) J
postParameterDecl	<i>pop</i>	FunctionDefTemplate	public void Foo((RType J)){ <statements; sep= '\n '> }
preCallStmt	<i>push</i>	CallStmtTemplate FunctionDefTemplate	Bar(<args; sep=', '>);
preVarNameExp	<i>push</i>	VarNameExpTemplate CallStmtTemplate FunctionDefTemplate	J
postVarNameExp	<i>pop</i>	CallStmtTemplate FunctionDefTemplate	Bar(J);
postCallStmt	<i>pop</i>	FunctionDefTemplate	public void Foo(RType J){ (Bar(J);) }

Fig. 9. A table illustrating the state of the translation template stack over the course of operation Foo's translation to Java.

information from C 's AST node into T as attributes. Finally, we push T onto the global template stack.

2. Over course of traversing C 's children, we modify the top of this stack arbitrarily—filling in any additional attributes. However, in the event that a child, C' , is as complicated as C , we perform the same steps as in (1), pushing a new template T' representing C' onto the stack, and walking its respective children.
3. In post visitor methods ($\text{post}C$), we simply pop the filled (completed) template off the stack, and either print it out (if we're done) or add it as an appropriate attribute to the new/current top of the stack.

We have found the tree walker to be an invaluable asset to this process, as not only do the pre and post operations for each construct allow us to produce arbitrarily complicated, nested blocks of structured output, but also provide a valuable context from which to instantiate, fill, and manipulate any relevant templates.

In summary, the only actual work being performed within the translator is forwarding information collected from individual AST nodes, to a series of externally defined templates. This allows us to exploit (in design pattern parlance) a strict model view controller (MVC) separation in the translator's codebase between the mechanism that does the AST visiting (controller), the individual AST nodes from which we're forwarding information to templates (model), and the external file containing all available C or Java language templates which shape our output (view) [12,13].

6 Results and Benefits

The dynamic tree walker is a unique and innovative reengineering approach that has yielded a number of benefits for our ongoing research project. The initial benefit was its rapid implementation. This allowed us to begin work on new versions of components with very little delay and with no interruption to the compiler's existing operation. In fact, it allowed for simultaneous operation of both new and old components—each sharing the same AST class hierarchy but using their distinct tree traversal mechanisms.

Another major benefit of the dynamic tree walker is that it adds an entirely new layer of abstraction to the visitor pattern. The visitor pattern already mandates the decoupling of the traversal logic from the visitation logic, but our design also separates the traversal logic from the structure of the tree itself. Because the structure of the tree is extracted from the code dynamically at runtime, changes can be made to the AST classes and the tree walker will seamlessly adjust. The visitors, on the other hand, may need to be adjusted in cases where an existing part of the tree was renamed or removed—though perhaps not when adding to the tree. This is due to the fact that visitation logic is, by necessity, coupled with the tree structure.

In many compiler compilers, the relationship between the traversal algorithm and the tree structure is established in a pre-runtime code generation phase. This

is usually accomplished by auto-generating AST classes with hard-coded traversal logic extracted from a grammar or other definition file. Our reengineering approach, however, does not require us to rewrite our legacy AST classes (or any part of them). Furthermore, it largely avoids the code generation step. It may be desirable to generate the abstract `TreeWalkerVisitor` class ahead of runtime to simplify the creation of new components by providing method declarations to override (indeed, we have opted for this approach). However, because methods are invoked using dynamically-constructed method names, this is not strictly necessary.

Finally, the dynamic tree walker is reusable and maintainable. As has been clearly demonstrated, there is very little coupling in our design. This allows the tree walker to be effortlessly reused for any number of components in our compiler. The maintainability is also enhanced by the fact that the traversal logic is contained within a single class rather than distributed over all AST classes. We have already leveraged this benefit by easily making additions and changes to the traversal such as inserting virtual nodes on-the-fly. Overall, the dynamic tree walker has saved a large amount of development time by eliminating the need to rewrite legacy code while also providing a rapidly implementable mechanism for continued iteration and evolution of the software.

7 Conclusion

Reengineering legacy code can be a time-consuming and often encumbering task for long-running software projects. There are usually no ideal solutions. Typical approaches carry stark tradeoffs—refactor the legacy code at a short-term development cost, or implement a quick work-around to simply defer the need for a permanent solution. In our case, where the problematic code involved a misapplied visitor pattern, we were able to leverage the reflective visitor pattern to minimize the drawbacks of the above alternative approaches. The reflection-based mechanism is a quickly implementable solution yielding long-term benefits for ongoing development, an approach we believe can be replicated for similar reengineering efforts. An extended code sample of our solution is provided in the [Appendix](#) for additional reference in reproducing our efforts.

8 Acknowledgements

This research is funded in part by US NSF grants CCF 1161916 and DUE 1022941. We acknowledge the contributions of the past and present members of the RESOLVE Software Research Groups (RSRG) at Clemson and Ohio State for the verifying compiler project.

References

1. Mai Y, de Champlain M. Reflective visitor pattern. *Technical Report*, Concordia University. URL http://hillside.net/europlop/HillsideEurope/Papers/EuroPLOP2001/2001_MaiEtAl_ReflectiveVisitorPattern.pdf.

2. Pati T, Hill JH. A survey report of enhancements to the visitor software design pattern. *Software: Practice and Experience* 2014; **44**(6):699–733, doi:10.1002/spe.2167. URL <http://dx.doi.org/10.1002/spe.2167>.
3. Bravenboer M, Visser E. Guiding visitors: Separating navigation from computation. *Technical Report*, Utrecht University 2001.
4. Sitaraman M, Adcock B, Avigad J, Bronish D, Bucci P, Frazier D, Friedman H, Harton H, Heym W, Kirschenbaum J, *et al.*. Building a push-button resolve verifier: Progress and challenges. *Formal Aspects of Computing* 2011; **23**(5):607–626, doi:10.1007/s00165-010-0154-3. URL <http://dx.doi.org/10.1007/s00165-010-0154-3>.
5. Welch D, Cook C, Sun YS, Sitaraman M. A web-integrated verifying compiler for resolve: A research perspective. *Proceedings of the 7th India Software Engineering Conference, ISEC '14*, ACM: New York, NY, USA, 2014; 12:1–12:6, doi:10.1145/2590748.2590760. URL <http://doi.acm.org/10.1145/2590748.2590760>.
6. Cook CT, Drachova S, Hallstrom JO, Hollingsworth JE, Jacobs DP, Krone J, Sitaraman M. A systematic approach to teaching abstraction and mathematical modeling. *Proceedings of the 17th ACM Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE '12*, ACM: New York, NY, USA, 2012; 357–362, doi:10.1145/2325296.2325378. URL <http://doi.acm.org/10.1145/2325296.2325378>.
7. Pati T, Hill JH. A survey report of enhancements to the visitor software design pattern. *Software: Practice and Experience* 2014; **44**(6):699–733, doi:10.1002/spe.2167.
8. Levine J, Mason T, Brown D. *Lex & Yacc, 2nd Edition*. Second edn., O'Reilly, 1992.
9. Gagnon E, Hendren L. Sablecc, an object-oriented compiler framework. *Technology of Object-Oriented Languages, 1998. TOOLS 26. Proceedings*, 1998; 140–154, doi:10.1109/TOOLS.1998.711009.
10. Gamma E, Helm R, Johnson R, Vlissides J. *Design Patterns: Elements of Reusable Object-oriented Software*. Addison-Wesley Longman Publishing Co., Inc.: Boston, MA, USA, 1995.
11. Parr T. A functional language for generating structured text (Draft). URL <http://www.cs.usfca.edu/~parrt/papers/ST.pdf> 2006; .
12. Parr TJ. Enforcing strict model-view separation in template engines. *Proceedings of the 13th International Conference on World Wide Web, WWW '04*, ACM: New York, NY, USA, 2004; 224–233, doi:10.1145/988672.988703.
13. Krasner GE, Pope ST. A cookbook for using the model-view controller user interface paradigm in smalltalk-80. *J. Object Oriented Program.* Aug 1988; **1**(3):26–49.

9 Appendix

Listing 1.4. Reflective Visitor Pattern Code

```
public abstract class TreeWalkerVisitor {
    public boolean walkAbstractFunctionExp(AbstractFunctionExp data) { return false; }
    public void preAbstractFunctionExp(AbstractFunctionExp data) {}
    public void midAbstractFunctionExp(AbstractFunctionExp node,
        ResolveConceptualElement previous, ResolveConceptualElement next) {}
    public void postAbstractFunctionExp(AbstractFunctionExp data) {}
    // Continue for all types of AST nodes...
}

public class TreeWalker {
    private TreeWalkerVisitor myVisitor;

    public void visit(ResolveConceptualElement e) {
        if (e != null) {
            if (!walkOverride(e)) {
                invokeVisitorMethods("pre", e);
                java.util.List<ResolveConceptualElement> children = e.getChildren();
                if (children.size() > 0) {
                    Iterator<ResolveConceptualElement> iter = children.iterator();
                    ResolveConceptualElement prevChild = null, nextChild = null;
                    while (iter.hasNext()) {
                        prevChild = nextChild;
                        nextChild = iter.next();
                        invokeVisitorMethods("mid", e, prevChild, nextChild);
                        visit(nextChild);
                    }
                    invokeVisitorMethods("mid", e, nextChild, null);
                }
                invokeVisitorMethods("post", e);
            }
        }
    }
}

public abstract class ResolveConceptualElement {
    public List<ResolveConceptualElement> getChildren() {
        Deque<Class<?>> hierarchy = new LinkedList<Class<?>>();
        Class<?> curClass = this.getClass();
        do {
            hierarchy.push(curClass);
            curClass = curClass.getSuperclass();
        } while (curClass != ResolveConceptualElement.class);
        List<ResolveConceptualElement> children = new List<ResolveConceptualElement>();
        ArrayList<Field> fields = new ArrayList<Field>();
        while (!hierarchy.isEmpty()) {
            curClass = hierarchy.pop();
            Field[] curFields = curClass.getDeclaredFields();
            for (int i = 0; i < curFields.length; ++i) {
                fields.add(curFields[i]);
            }
            curClass = curClass.getSuperclass();
        }
        Iterator<Field> iterFields = fields.iterator();
        while (iterFields.hasNext()) {
            Field curField = iterFields.next();
            if (!Modifier.isStatic(curField.getModifiers())) {
                curField.setAccessible(true);
                Class<?> fieldType = curField.getType();
                if (ResolveConceptualElement.class.isAssignableFrom(fieldType)) {
                    children.add(ResolveConceptualElement.class.cast(curField.get(this)));
                } else if (java.util.List.class.isAssignableFrom(fieldType)) {
                    Class<?> listOf = (Class<?>)((ParameterizedType) curField.getGenericType()).getActualTypeArguments()[0];
                    java.util.List<?> fieldList = java.util.List.class.cast(curField.get(this));
                    if (fieldList != null && fieldList.size() > 0
                        && ResolveConceptualElement.class.isAssignableFrom(listOf)) {
                        children.add(new VirtualListNode(this, curField.getName(),
                            (java.util.List<ResolveConceptualElement>) fieldList,
                            (Class<?>)((ParameterizedType) curField.getGenericType()).getActualTypeArguments()[0]));
                    }
                }
            }
        }
        return children;
    }
}
```