Semantic Parsing of Java I/O in Novice Programs for an Online Grading System

William Madden

Follow this and additional works at: https://digitalcommons.montclair.edu/etd

Part of the Computer Sciences Commons
Semantic Parsing of Java I/O in Novice Programs for an Online Grading System

Abstract

Beginning programming students have access to sophisticated development tools that enable them to write syntactically correct code in a straightforward manner. However, code that compiles and runs can still execute poorly, or with unintended results. We present a tool, based on an open-source parser-generation product written in Java, that performs semantic analysis of novice Java code. Specifically, the present investigation concerns the semantics of Java output methods, particularly when they are enclosed within iterative structures in the language. The effort will be to guard against threats that such methods pose to system integrity and performance, intercepting them prior to runtime. The approach used here closely models the analysis a human reviewer would perform, given a printed copy of the code. The tool is an open-source product, like the parser generator, and is also written in Java. As such, it is written to be extensible. The tool will be integrated into a larger research project underway at Montclair State University which involves the development of an online grading system for students in beginning computer programming courses.
Semantic Parsing of Java I/O in Novice Programs for an Online Grading System

by

William Madden

A Master’s Thesis Submitted to the Faculty of

Montclair State University

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

May 20, 2005
SEMANTIC PARSING OF JAVA I/O IN NOVICE PROGRAMS
FOR AN ONLINE GRADING SYSTEM

by
WILLIAM MADDEN

A THESIS
Submitted in partial fulfillment of the requirements
for the degree of Master of Science in
the Department of Computer Science in
the Graduate Program of
Montclair State University
May 2005
Acknowledgements

I would like to thank my wife, Carol Heinz, and my son, Kevin Madden, for their endless reserves of support and patient understanding while I worked late into too many evenings. I would like to thank my extended family, including my father, Ward Madden, my brother, Robert Madden, and my sister, Sara Madden and their respective families, for words of encouragement all along the way. Thanks especially go to Dr. John Jenq, my thesis sponsor who, all along, provided much-needed support and encouragement to see this project through to its conclusion, and who taught best by asking the right kinds of questions. My deep gratitude too, to Dr. Dorothy Deremer and Dr. Edward Boyno who, together with Dr. Jenq, served on the thesis committee for this work. Thanks to Dr. James Benham, the graduate student advisor, for suggesting that I consider doing original research in the first place and who provided kind words of encouragement throughout. A debt of gratitude goes out to the open-source community, and to Terence Parr in particular, for supporting the work that makes the antlr parser-generating program such a valuable tool. Thanks, too, go to my peers at Bergen Community College who were always curious to know how the research was coming along and who were always helpful with kind words and praise; and to the Bergen Community College administration who provided generous and much needed tuition assistance to make this possible. And, finally, thanks go to my fellow graduate and undergraduate students at Montclair, as well as my own students at Bergen Community College, who must find it amusing, to say the least, that I am both teacher and student. Humble thanks to all.
## Table of Contents

Table of Figures ............................................................................................ vii

Chapter 1: Introduction ................................................................................. 1
  1.1: Overview .......................................................................................... 1
  1.2: Overall Project Description .............................................................. 4
  1.3: Research Topic and Rationale ......................................................... 4

Chapter 2: Literature Review ..................................................................... 6
  2.1: Background and Related Work ....................................................... 6
  2.2: Literature Review ........................................................................... 6

Chapter 3: Methodology and Implementation ............................................. 12
  3.1: Methodology ................................................................................. 12
  3.2: Implementation ............................................................................ 13

Chapter 4: Results and Conclusions .......................................................... 22
  4.1: Results .......................................................................................... 22
  4.2: Conclusions ............................................................................... 27

Chapter 5: Future Work ............................................................................. 29

References .................................................................................................... 31

Appendix A: Java 1.4.2 Input/Output Classes ........................................... 35

Appendix B: List of Java 1.4.2 Output Methods ....................................... 37

Appendix C: Java 1.4.2 Output Classes and Associated Output Methods .. 38

Appendix D: List of Java 1.4.2 tokens ...................................................... 46

Appendix E: for, while and do structures tested with IOScan .................. 48

Appendix F: Source Code listing for IOScan.java ................................. 49

Appendix G: Resources for Information on Semantic Parsing .................. 63
Table of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>16</td>
</tr>
<tr>
<td>Figure 2</td>
<td>18</td>
</tr>
<tr>
<td>Figure 3</td>
<td>24</td>
</tr>
<tr>
<td>Figure 4</td>
<td>25</td>
</tr>
<tr>
<td>Figure 5</td>
<td>25</td>
</tr>
<tr>
<td>Figure 6</td>
<td>26</td>
</tr>
<tr>
<td>Figure 7</td>
<td>26</td>
</tr>
<tr>
<td>Figure 8</td>
<td>27</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1: Overview

The study of languages is the province of linguistics. One of the notable researchers in the field, still active as of this writing, is Noam Chomsky, whose work has influenced not only the general field of linguistics itself, but has found fundamental application in computer science as well. His early description of a hierarchy of language grammars [4] has provided the basis for describing and formalizing most computer languages in use today. Most computer languages are described as implementations of context-free grammars (Type 2 in Chomsky’s now-famous hierarchy). A widely-used tool for notating context-free grammars is Backus-Naur Form (BNF), developed in the late 1950s, immediately after Chomsky’s seminal work. John Backus, who had already invented FORTRAN, the first widely-used high-level computer language, initially created what came to be known as BNF while working on the development of ALGOL, itself a very influential language. Peter Naur also contributed significantly to the development of BNF.

Inherent in the study of computer languages are notions such as grammar, syntax, and semantics. Loosely, grammar refers to the formal set of rules that results in the production of language elements and how they may relate to one another. Syntax refers to the sequencing of language elements to form language expressions. Semantics refers to the meaning of a language expression. Current development environments for computer languages provide complete implementations of the grammar of the language, a full set of tools for suggesting grammatical and syntactic elements to programmers, and fairly sophisticated tools for debugging grammatical and syntactic errors in programs.
during the development process. All development environments also provide feedback to programmers during the compilation process, particularly with regard to syntactical errors. Thus programmers can fairly easily develop programs that compile successfully—that is, are grammatically and syntactically correct. However, the semantic structures they have created (that is, the meaning or intended purpose of the program) may still be found wanting. Most development environments provide only rudimentary feedback for runtime errors, and it is often the case that the error reported is not the error that caused the problem. This is often misleading to novice programmers who can spend a frustrating amount of time trying to apply the wrong corrections to a problem. Run-time errors are often the result of faults in the programmer’s logic, or in the programmer’s assumptions about possible run-time behaviors while developing the program.

A more sophisticated semantic analysis of source code begins with the ability to resolve code down to individual syntactic elements. Tools commonly used for this purpose are known as lexical analyzers (lexers for short, known more generally as language recognizers) and parsers. A lexer reads a stream of characters (source code in this case) and, acting as a finite state accepter, produces a set of lexemes, or acceptable strings of characters. These strings are then evaluated to produce a set of lexical tokens that can then be processed by a parser. Parsers are programs that, given a particular formal grammar (most often notated in BNF), analyze a stream of tokens and build a data structure, usually a tree or a graph that represents the grammatical relationships among the syntactic elements. This structure can become the basis for a semantic analysis of the code.
In practice, parsers are usually the result of parser generation programs. This is because a particular formal grammar is an arbitrary construct and often subject to change and elaboration. Each time the grammar changes, a new parser must be generated. Most so-called parsers today are actually parser generators and are dependent on accurate, current grammars. In the case of Java, which is still a relatively young language, the grammar is in a state of flux.

There are two basic approaches to parsing code: top-down and bottom-up [24]. Bottom-up parsers are typically used in the compilation process. They are characterized as LR parsers, meaning that source code is read from left to right and that syntactic elements are constructed or derived from the right-hand end of a particular input string. LR parsers are “Left-to-right, Right-hand derivation” parsers. Yacc, a common parser found on UNIX systems, is one example of an LR parser. Top-down (or LL) parsers also read input from left to right, but use left-hand derivation. The distinction is important because the two approaches produce syntactic elements in a different order from one another. This also has implications for parser performance and for the kinds of grammars for which the particular approach is suitable.

A particular type of LL parser is the $LL(k)$ recursive descent parser. Such a parser “looks ahead” $k$ tokens to make parsing decisions. The ability to examine tokens before processing them is necessary in practice because modern programming languages, such as Java, are not purely context-free. Recursive descent parsing is particularly suitable for the production of Abstract Syntax Trees, which are explained more fully below (section 3.2). The particular parser used in this investigation, antlr, is an $LL(k)$ recursive descent parser.
1.2: Overall Project Description

This research is part of a larger project already underway at Montclair State University. An online system is being developed that will provide feedback to students in two introductory undergraduate programming courses, regarding their Java coding exercises. The purpose of this larger project is to create an expert system that analyzes the semantics of Java code and provides useful feedback to students as they develop their coding skills. The larger research effort is being directed by Dr. John Jenq, Computer Science Department, Montclair State University.

In developing expert systems, concerns arise regarding threats posed to system integrity by I/O operations, particularly as a result of poorly-written or even malicious code. For example, the author wrote a simple program that wrote “Hello, World” to the hard drive on a 2.4 GHz/Pentium 4 computer with 512 MB of RAM (a very typical student machine at the time of this writing). The only unusual feature of the program was that the write to disk occurred inside an infinite loop. In 6 seconds of execution the program created a 400 MB text file. The operating system attempted to buffer the file in memory; however, there was not enough available primary memory and it began using virtual memory. The result was so deleterious to performance that it was quicker to reboot the machine than to wait for it to stabilize itself.

1.3: Research Topic and Rationale

The present investigation concerns the semantics of Java output methods, particularly when they are enclosed within iterative structures in the language. The effort will be to guard against threats that such methods pose to system integrity and
performance, intercepting them prior to runtime, specifically to avoid scenarios such as the one described above.

An underlying assumption is that the code to be examined (referred to as novice code) has already been compiled and that the development environment has reported no compilation errors, that is, the code is assumed to be syntactically correct. Today, most programming students have access to sophisticated development environments that do a lot of the syntax-checking as the program is being written and, in fact, provide drop-down lists of suggested keywords, classes, methods and arguments as the student types out the code. For this reason, it is becoming a relatively straightforward process for novices to be able to produce syntactically correct code. However, much in the same way that correct spelling of all the words in an English sentence does not ensure a grammatically correct sentence and, in turn, a grammatically correct sentence does not ensure semantic correctness (an intended meaning), so too with source code. A novice can write correct code that will execute, but it may execute poorly or with unintended results. A semantic analysis of the code can yield important feedback to both students and instructors prior to run-time and forestall errors or potentially harmful run-time behaviors.
Chapter 2: Literature Review

2.1: Background and Related Work

The author identified four areas of related research in the literature: semantic analysis of Java, handling malicious code, intelligent tutoring systems, and common errors in novice code.

The broadest of these research areas, semantic analysis, finds wide application: compiler generation, software testing, software quality assurance, program analysis and verification, exception handling, class analysis, software engineering tool development, code optimization, and constraint-based program analysis, to name a few. Much of the research in semantic analysis focuses on the technique of performing static analysis of code, that is, analyzing the semantic structure of source code without executing it. That is the approach used here.

2.2: Literature Review

The literature regarding semantic analysis revealed several alternative approaches to the development of automated grading systems that may be useful for the development of the overall project. A number of projects describe visualization tools that help novices to more easily discern structural and semantic relationships within their code [2] [3] [14]; however, these do not provide specific feedback mechanisms concerning specific semantic issues in code. Huynh [10], Fabry [7] employ variations of a pattern-matching approach; that is, source code is compared to a predefined ‘template’ program to identify syntactic and/or semantic differences. This may be a very worthwhile approach, especially with first-semester programming projects. The approach used in this
investigation is more open-ended and does not rely on modeling ideal code solutions. A particularly interesting approach [17] [19], is to use the structured data capabilities of XML in the semantic parsing process. One reason for this is to be able to take advantage of the rich array of tools available for XML that provide powerful ways to analyze semantic structures.

Much of the interest in identifying and handling malicious code is fueled by the vigorous interest in security issues these days, although there is an offshoot of this research that deals specifically with the phenomenon of run-time termination on mobile devices. This research often describes particular tools and or techniques. Rabek [20] describes a tool that performs a static semantic analysis of code, identifying Win32 API system calls. The tool then monitors the system calls made at runtime, verifying that they conform to the calls determined by the semantic analysis. The first half of this approach (the semantic analysis) is similar in principle to the approach used here, though it is not language specific and it is expressly designed to work only on Windows platforms. The approach used here also does not do any runtime monitoring – all analysis and feedback is provided during the static analysis.

Research in intelligent tutoring systems or agents (the term *de jour*) extends back at least into the 1960s, with a host of approaches and techniques that have evolved and changed as technology has grown and access to computers and the internet has grown to near ubiquity, at least in industrialized countries. A spin-off of this research has delved into the teaching and learning of programming languages, particularly regarding automated tools that can help teachers evaluate code written by novice programmers. Six specifically address the teaching of Java [23] [22] [5] [6] [15] [8].
Sykes [22] describes a prototype for a system that will involve a small subset of the Java language. This may be an approach worth investigating, again, especially for first-semester student programming projects. It remains to be seen, once the prototype is expanded to include the whole language how closely it will resemble the author’s own efforts. The present investigation seeks to include the whole language from the outset.

DePasquale [5] also employs an approach that works with a subset of the full language; it is interesting to note that he does not find significant differences in performance between students who begin by learning subsets of the language in simplified development environments compared to students who learn from the outset using the full language in complete development environments. He does note that students using the simplified environments were more satisfied with their experience than students in the traditional environment.

Truong [23] begins by identifying a number of common semantic errors made by novice programmers (this was done through his own extensive literature review and then confirmed empirically on his own campus). He employs software metrics and structural metrics to analyze and evaluate code semantics, based on a comparison of semantic parse trees of student code with model code. The code assignments are “fill-the-gap” in nature, i.e., a skeleton structure is supplied and the student must complete what are called “well-formed gaps” in the code so the result will compile correctly. This approach seems particularly well-suited for first semester students and might be worthy of investigation for use within the larger research project at the University.

Truong employs metrics that quantitatively measure how far a novice’s code departs from predefined stylistic guidelines or from samples of model solution code. An
interesting side note here is that Badros [1] provides a very nice set of examples that implement these kinds of metrics using XML-wrapped java source code.

The approach that Truong and Badros employ, providing a ‘template’ to which the novice’s efforts are compared, are more recent implementations of what appears to be a ‘pattern-matching’ approach that has been used most often historically. A number of earlier efforts were investigated, more to understand the evolution of the tools, than to gain insight into particular approaches. An approach used by Jackson [11] seems representative: a tool that uses yacc and lex to compare student code to a ‘correct’ solution. Huynh and Fabry, noted earlier, seem also to have refined and elaborated on this historical approach.

Kumar [15] has created a number of tutors for use with either Java or C++ that address specific topics in beginning programming (the specific reference given here deals with expression evaluation). The tutor produces snippets of code that students must then evaluate. In more extended cases, the evaluation is done interactively in a step-wise fashion, essentially, tracing the execution logic of the code.

Hristova [8] has developed a tool, named Expresso, that specifically concerns itself with providing pre-runtime feedback for a set of common Java programming errors. Part of her research, like Truong, produced a list of common syntax, semantic and logic errors in Java. It appears that most, if not all, of the errors listed are actually caught by most current compilers. Part of the reason for the development of her tool was to help students interpret often cryptic compiler error message that can often actually be misleading in terms of the necessary corrections.
Etheridge [6] provides a very interesting tool called CMeRun that pre-processes source code, providing a snap-shot of the condition of variables and various kinds of output as each line of code executes. It provides something akin to the step-through execution that most debuggers produce within current IDEs. It allows students to step through loops providing the opportunity to break out at each iteration.

There are also other tools that provide visualization tools and various other techniques for helping novice programmers better understand what they are doing. Kumar already provides a concise summary of these approaches, including Hristova and Etheridge in his summary.

Like Hristova and Etheridge, and in contrast to Kumar, Truong and Sykes, the author’s approach is designed to work with non-prescribed projects. The present approach works with any Java code and does not expect the code to match any pre-existing or closed solutions. This approach seeks to determine runtime results prior to runtime and prevent syntactically correct but troublesome code from executing. In this regard, the concerns in the present investigation are somewhat closer to Reiss [21] although his interest is in developing CASE tools for large-scale enterprise projects. Reiss describes a process whereby semantic parsing is largely employed to maintain a symbol table and in which the symbol table is continuously updated during incremental parsing, effectively generating a model of runtime behavior. This will become an important feature in future extensions of the current investigation. Lapierre [16] also describes a commercial tool developed at Bell Labs Canada, designed for large-scale projects, that builds an Abstract Semantics Graph as its central tool for analyzing source code. Since it is proprietary, details regarding its implementation are not available.
Finally, there is a significant body of research that identifies common novice errors in syntax and logic. Truong and Hristova both investigated a number of common programming issues specifically related to Java, drawing upon their own experiences and generating surveys of their peers, as well as reviewing the literature. Their work will be very useful in determining future directions for this research. As far as the author knows, at this writing, there have not been any investigations into the semantic issues involved in generating I/O from within iterative structures in Java.

Moreover, the author’s approach appears somewhat unique in that it attempts to model the behavior of a human reviewer. A human reviewer could, for instance, read the following code: 

```java
for (x=0; x<10; x *=1);
```

and can model its runtime behavior without actually compiling and running the code. After performing the virtual run, the human reviewer can predict runtime performance and provide feedback to the novice programmer, pointing out, in the example used here, that while the loop will compile and execute, it is infinite. That is essentially the approach used here. The author has previously reported preliminary findings at an earlier stage in this research [18].
Chapter 3: Methodology and Implementation

3.1: Methodology

There are a host of open-source tools and resources readily available that provide a foundation for doing semantic analysis of Java code. The present effort seeks to build an open-source tool, based on existing tools and resources, that can anticipate runtime errors, saving both students and instructors time and effort, and helping to maintain system integrity.

Complete semantic analysis in any language is extremely difficult and, ultimately, open to much interpretation and debate. Therefore, a number of important assumptions and constraints were imposed in the present case in order to define a meaningful scope for this research and to provide an extensible basis for future research:

(1) The tools described in this investigation work with Sun Microsystems' official Java 1.3.1 specification.

(2) The tools are designed to examine Java source code produced by students in beginning computer science courses. Such code will hereinafter be referred to as 'novice code' or 'novice programs'.

(3) It is assumed that the novice code under examination has first been successfully compiled, that is to say, the code is syntactically correct.

(4) For the purposes of this investigation, the tools perform a semantic analysis of novice code with the following constraints:

   a. All current Java (1.4.2) output classes and methods are included in the investigation. Input classes and methods are not part of the present investigation.
b. The tools perform semantic analysis of output classes and methods only when they occur within iterative structures (specifically, *for*, *while* and *do* loops). Invocations of output classes and methods that execute singly in-line are not investigated.

c. Only numeric loops are considered. That is, initial conditions for entrance into a loop, exit conditions from the loop and the loop iterator are assumed to be numeric expressions or assignments. Loop controls using other means (boolean expressions, character or string expressions, pointers and so on) may be the subject of future investigations.

d. Only loop control identifiers that have no dependencies are considered in the present investigation. Typically, it is assumed that loop control identifiers are declared and initial values assigned early in the code with no further modification before their use in the loop. More complex or subtle manipulation of loop control identifiers is not within the scope of the present investigation.

e. Only simple 'in-line' iterative structures are considered. Nested loops, embedded method calls, recursive structures, and so on, may be the subject of future investigations.

3.2: Implementation

This research began with an investigation of the Java language specification itself. A BNF grammar for version 1.3.1 was chosen because it was the most recent available for the language at the time the core of this research was conducted. As of March, 2005 there is now a BNF grammar available for the most recent version of the language
(version 1.4.2). It is anticipated that, at most, only minor changes will be needed to enable the author's tools to work with the newer grammar. In fact the author has used the 1.3.1 grammar with the complete set of 1.4.2 output classes and methods with no ill effect.

An examination of the official language specification on Sun Microsystems' web site [12] revealed that Java 1.4.2 contains 149 output classes. Because of polymorphism, these classes share 57 output methods. Java I/O classes are listed in Appendix A. Java output methods are listed in Appendix B. Appendix C lists all Java output classes, together with their associated methods.

Next, a language parsing engine was needed. A number of open-source products are available for this purpose: antlr, yacc, Bison, Semantic, lex, and flex among others. See Appendix G for links to more information concerning a number of these alternatives. There are also commercial products available that the author did not choose to investigate.

Antlr (ANother Tool for Language Recognition) [9] was chosen for the present investigation. It is the only parser that has a complete open-source implementation in Java. This was important since the author's own tools are also built using Java. It is also an extremely fully-featured product and is constantly growing and developing thanks to its large community of users and developers (approximately 45,000). It has also been ported to UNIX, Linux and Windows platforms. The antlr web site provides complete BNF grammars to a number of languages besides Java including C++, C, Ada, HTML, Python, Oracle SQL, and C#. The source code for antlr is currently written in Java, but there is a C++ version as well.
Antlr contains a parser generator; that is, it is implemented by first supplying a grammar in BNF notation. It is straightforward one-time process to build the actual parsing tools. Once this is done for any arbitrary language, antlr can parse source code in that language and produce a data structure known as an Abstract Syntax Tree (AST), to represent the source code. Antlr also contains a rich set of classes and methods for working with ASTs, token streams, tokens and lexemes.

The reader may wish to refer to Figure 1 for the discussion in the next four paragraphs. It shows a simple Java program and the AST that it produces.

An Abstract Syntax Tree is a recursive tree structure and is the result of recursive descent LL(k) parsing. LL(k) parsing is particularly suitable for Java grammar which is not purely context-free. Recursive descent parsing is helpful because the resulting structure closely resembles the structure of the code that produced the AST.
Figure 1: An antlr-produced Abstract Syntax Tree (AST) and the source code that produced the tree.
An AST for a standalone complete Java program is structured as follows: The AST root has a FirstChild: the CLASS_DEF node. The CLASS-DEF node is composed of a FirstChild and four siblings: an internal MODIFIERS node, three leaf nodes (the Class name, an EXTENDS clause and an IMPLEMENTS clause) and an internal OBJBLOCK node. The last of these, the OBJBLOCK, contains at least one internal METHOD_DEF node (main) and could contain Class-level declarations and other METHOD_DEF nodes as well. The METHOD_DEF node contains four internal nodes: MODIFIERS, TYPE, PARAMETERS, and SLIST. The last of these, SLIST, finally contains the code body.

Any internal node has a FirstChild node at least. There may be other child nodes; however, since antlr is a recursive descent parser (described below) these are only accessible in a serial fashion as getNextSibling methods relative to the FirstChild node and are not directly accessible from the parent node. Many, but not all, internal nodes are binary; that is, they have a FirstChild and a NextSibling, or in more standard parlance, a LeftChild and a RightChild. However, there are a number of internal nodes (chiefly SLIST and ELIST nodes) that may have any number of child nodes (or, more precisely, a FirstChild node and several NextSibling nodes). A leaf node, by definition, has no child nodes, but may have NextSibling nodes. These NextSiblings may be either internal nodes or leaf nodes. There are no methods built into antlr that permit backtracking (except as the inevitable result of ascending out of a recursion), so the concept of a parent node is absent.

Antlr implements recursive descent parsing, which means, among other things, that 'walking' an AST involves traveling forward-only through source code in a recursive
depth-first manner. The algorithm uses a `getFirstChild` method to locate FirstChild nodes first. Then, if there are no more depth-first children, it uses a `getNextSibling` method to identify any sibling nodes, ascending out of the recursion along the way.

The general pseudocode for a recursive descent is shown in Figure 2.

```java
1 boolean searchSubtree(AST tree, String searchItem) {
2    AST thisNode, child
3    boolean itemFound = false
4    thisNode = tree
5    while thisNode is not null {
6        child = thisNode.getFirstChild
7        if child is not null
8            ifFound = searchSubtree(child, searchItem)
9            if thisNode contains searchItem
10               itemFound = true
11               thisNode = thisNode.getNextSibling
12        }
13    return itemFound;
14 }
```

Figure 2

Line 8 contains the recursive statement and lines 6-8, contained in the while loop, illustrate the depth-first nature of the recursion. Line 11, also contained in the while loop, illustrates the serial nature of the 'walk' among sibling nodes. To process the last of several siblings, each of the previous siblings must first be processed in order before moving to the last sibling.

EXPR nodes are of particular importance. As simple as the code behind the example shown in Figure 1 is, its AST contains 5 EXPR nodes, used variously to parse a simple assignment statement, to express the exit condition for a while loop, to hold a println method call, to hold a literal string, and to contain the iterator expression. Because EXPR nodes have such varied uses, their internal structure (and nesting)
requires careful attention. One phase of this research addressed the development of Java
code to evaluate EXPR nodes.

The basic technique in this investigation was to develop a software tool, called
IOScan, that calls antlr methods to build an AST, walks the tree searching for iterative
structures such as for, while and do, and then locates any Java output methods possibly
executing within those structures. Any loops containing output methods are examined to
see whether or not they are well-formed. In this context, an expression such as: 
\textit{for (x=1; x<=10; x++)}, is well-formed, whereas: 
\textit{for (x=1; x<10; x--)}, is mal-formed. For a
complete list of iterative structures (both well-formed and mal-formed) tested in this
investigation see Appendix E. The complete source listing for IOScan.java is available
in Appendix F.

For performance reasons, the search for output methods is conducted by
comparing nodes to a hashed list of Java's 57 output methods. Iterative keywords for, 
while and do were also hashed, though it is doubtful such an effort was really necessary,
at least in regard to performance. (At one point the author had hypothesized the need for
a larger list of keywords, but the scope of this research effort is restricted to just for, 
while and do).

The IOScan logic addresses conditions where loop-control identifiers are numeric
and have no dependencies. In this case, for loops in novice code are generally very
tractable. The parenthetical expression following the keyword for defines entry, exit and
iterator expressions for the loop-control identifier and these are easily located in a
recursive descent search. While and do loops are not so straightforward. The only
structure immediately tied to the keywords while or do is the exit condition of the loop-
control identifier and, even then, the identifier's type is not known. It is necessary, because of the recursive descent nature of the parser, to do separate 'walks' through the AST to locate the declaration and initial condition of the loop control identifier, and the iterator expression.

A further problem to be solved involves the evaluation of loop-control identifier expressions. A great deal of time was spent investigating the development of a Runtime Identifier Model (RIM). The purpose of the RIM is twofold. It will provide:

(1) a symbol table which stores identifier names, along with their addresses, types and values (or expression addresses), and

(2) an expression evaluator which determines, through a modeling of run-time behavior, whether a particular identifier can be evaluated (especially if it is a loop-control identifier) and, if so, what it's value is at any given point in the code.

The author was able to develop symbol table structures easily enough, but developing an extensible expression evaluator proved to be very time consuming and needs to be the subject of future research. There are 163 possible tokenizable syntactic structures (including keywords, operators and symbols) that make up Java 1.4.2 (see Appendix D). A great many of these can find their way into EXPR nodes. An expression evaluator would need code for each token, to be able to evaluate it in context, essentially requiring something of the complexity of an interpreter for the entire language.

The development of a RIM, with a workable expression evaluator, is an important next step because such a tool could accommodate more subtle, complex and real-world uses of loop-control identifiers, particularly as novice programmers become more skilled.
Currently, IOScan examines the following kinds of expressions (in the examples supplied below n can be any valid numeric value, depending on identifier type).

Loop-control identifier initial condition:

Declarations of loop-control identifiers without assignment: \textit{int} \texttt{x};
Declarations of loop-control identifiers with assignment: \textit{int} \texttt{x} = n;
Simple assignment statements following earlier declarations: \texttt{x} = n;

Loop-control identifier exit condition:

Standard boolean comparisons: \texttt{x < n; x > n; x = = n; x <= n; x >= n; x ! = n;}
Loop-control iterator. Expressions in any of the following forms:

\texttt{x++; x--; x += n; x -= n; x *= n; x /= n; x = x + n; x = x - n; x = x * n; x = x / n;}
Chapter 4: Results and Conclusions

4.1: Results

In its present state of development, IOScan can evaluate simple in-line (un-nested) for, while and do loops using numeric loop-control identifiers that have no dependencies. Despite its rudimentary scope, there are still quite a number of unusual (but entirely possible) novice programs that can be written, which will compile and run, but which are still mal-formed and will perform either unpredictably, poorly, not at all, or will loop indefinitely.

Given the restrictions on expressions noted above, the following kinds of loop behaviors are possible (all examples are illustrated using for-loops for brevity, but while and do loops can be written that will behave identically).

(1) Well-formed loops. These are loops where the relationships among the initial condition, the exit condition and the iterator are such that the loop will execute a finite and predictable number of times. Example: for (x=1; x<10; x++)

(2) Infinite loops. Examples: for (x=1; x<10; x += 0) and for (x= -10; x<l; x /= 2)

(3) Loops where the iterator heads in the wrong direction. The following example is actually an infinite loop because, in Sun's implementation of Java, int-types are 2s-complement values and, instead of going out of range, will increment or decrement 'around to the other side' of their permissible range. Example: for (x=1; x<10; x--)

(4) Loops where the exit condition is the inverse of what it should be. This occurs when the novice programmer confuses > and < symbols. Example: for (x=1; x>10; x++)
(5) Loops where the iterator converges to 0 from either side (positive or negative). This may be intentional in some cases. Examples: \textit{for} \((x=1; \ x<10; \ x*=0.5)\) probably not intentional; \(\textit{for} \ (x=-10; \ x<-1; \ x*=0.5)\) perhaps intentional.

(6) Loops where the iterator converges to 0 alternating between positive and negative values. This is likely not done intentionally. Example: \textit{for} \((x=1; \ x<10; \ x*=-0.5)\)

(7) Loops where the iterator alternately diverges toward negative and positive infinity with each iteration. This is likely unintended. Example: \textit{for} \((x=1; \ x<10; \ x/=-0.5)\)

(8) Loops that converge to 0 in 1 iteration. This is a special case and likely not an intentional effort. Example: \textit{for} \((x=1; \ x<10; \ x*=0)\)

(9) Loops that produce an undefined condition. This is also a special case and likely not an intentional effort. Example: \textit{for} \((x=1; \ x<10; \ x/=0)\)

(10) Loops that use \(!=\) as the boolean operator in an exit condition. This may be intentional or otherwise. Example: \textit{for} \((x=1; \ x!=10; \ x=x+2)\)

(11) Loops where initial and exit conditions are the same. This is not likely an intentional effort. Example: \textit{for} \((x=1; \ x<=1; \ x++)\)

(12) Loops that iterate between the same negative and positive values. Not likely an intentional effort. Example: \textit{for} \((x=1; \ x<10; \ x/=1)\)

(13) Loops that iterate excessively. This might be intentional or otherwise. Currently \textit{IOScan} reports any loop governed by an arithmetically increasing or decreasing iterator that executes more than 1 million times. This is an arbitrary choice and can easily be made adjustable. Example: \textit{for} \((x=-1; \ x>-10000000; \ x=-2)\)
IOScan distinguishes among all the behaviors listed above. There can, of course, be ambiguous scenarios. For instance, it is quite possible in example (3) above that the novice programmer meant to write \( \text{for} (x=1; x \geq 10; x--) \), so that the error is not in the iterator expression, but in the exit condition expression. IOScan only seeks to distinguish between loops that execute in a well-behaved fashion (a finite, predictable number of times for numeric loop-control identifiers), and those that do not, or those that behave in very unconventional ways.

Sample runs for several novice programs are presented below (Figures 3 – 8). For a complete listing of feedback provided by IOScan, see the \texttt{feedback()} method in the source code listing in Appendix F.

```java
public class SimpleFor {
    public static void main(String[] args) {
        int x;
        for (x = 0; x <= 100000; x++) {
            System.out.println("Hello");
        }
    }
}
```

Figure 3: Properly formed for loop with println output method embedded
```java
public class SimpleFor {
    public static void main(String[] args) {
        int x;
        for (x = 0; x <= 1000000; x--) {
            System.out.println("Hello");
        }
    }
}
```

Figure 4: Mal-formed loop; println embedded; iterator should be ++; excessive iterations

```java
public class SimpleWhile {
    public static void main(String[] args) {
        int x;
        x = 1;
        while (x <= 100000) {
            System.out.print("Hello");
            x *= -1.5;
        }
    }
}
```

Figure 5: Mal-formed while loop; print method embedded; unusual iterator behavior
```java
public class SimpleWhile {
    public static void main(String[] args) {
        double x;
        x = 1;
        while (x <= 100000) {
            System.out.print("Hello");
            x = x / 1.5;
        }
    }
}
```

Figure 6: Mal-formed while loop; print method embedded; unusual iterator behavior

```java
public class SimpleWhile {
    public static void main(String[] args) {
        double x;
        x = 1;
        while (x <= 100) {
            System.out.println("Hello");
            x = x - -1.5;
        }
    }
}
```

Figure 7: An unusual iterator that actually works
```java
public class SimpleWhile {
    public static void main(String[] args) {
        double x;
        x = 1;
        while (x >= 100) {
            System.out.println("Hello");
            x = x - 1.5;
        }
    }
}
```

**Figure 8:** same as Figure 5 except exit condition is inverted

**4.2: Conclusions**

The particular concern of this investigation was to be able to produce a tool that can examine novice code and detect possible semantic issues resulting in the execution of code that could compromise host system integrity. The particular issue was the detection of I/O methods executing within iterative structures in Java. *IOScan* is able to do that, at least for iterative structures in which the loop control identifier has no dependencies, is initially declared and/or defined with simple assignment statements, the exit condition is a simple single Boolean comparison, and where the iterator is a simple single increment, decrement, multiplier or divisor statement. The program carries out an automated semantic analysis of Java programs and is built to be extensible. Its *modus operandus* is to emulate the behavior of a human reviewer, carrying out the same kind of pre-run-time analysis and predictive modeling of run-time behavior that a human reviewer carries out. Writing code that can automatically evaluate a wide range of novice programs and
provide useful feedback for a range of semantic issues in novice code is a large task. The present effort represents a rudimentary first effort in that direction.
Chapter 5: Future Work

Development of a successful runtime identifier model would be an important next step in extending the capability of IOScan. An important feature of such a model would be for it to be able to evaluate EXPR nodes (expressions) in an AST. The model suggested by Reiss (mentioned earlier [21]) could provide a useful starting point for development in this direction. An important issue to address is to be able to recognize identifier scoping and external dependencies such as method calls or user input at runtime. A next logical step would be to extend the runtime identifier model so that expressions involving non-numeric loop-control identifiers could be evaluated. Other directions could include the evaluation of nested loops and recursive structures.

There are also other semantic issues in novice code apart from the dangers of iterative output. Truong [23:319] mentions a number of these: improperly structured switch blocks, hard-coding of literals, poor scoping of identifiers, and unused identifiers. These could also be fruitful avenues for exploration and extension of the current research. Hristova [8] also describes a number of specific syntactic, semantic and logical errors throughout her article.

For the larger project of which this investigation is a part, it remains to be seen how the present research can be integrated into the system as it evolves. In addition to the approach described in this investigation, some of the approaches outlined in the literature review are worthy of investigation, particularly for first-semester students. Two approaches especially seem appropriate in this regard: starting novices with a simpler subset of the language making use of a relatively simple and uncluttered development environment, and providing projects of a “fill-the-gap-with-well-formed-code” nature.
There is no reason that either of these approaches could not be integrated with the author's own approach as the system is refined and developed.

A number of more recent approaches to the problem of providing intelligent tutoring systems rely on web-based delivery and XML-based parsing tools that have been developed. These approaches represent quite a different approach to providing novice programmers with useful feedback than what has been considered here and should be examined. Finally, more recent approaches are incorporating visualization tools for providing intuitive feedback to novices, taking advantage of the great strides that have been made in multimedia development in recent years. These too would be worthy of investigation.
References


[9] http://www.antlr.org. This is the antlr web site. The open-source product can be freely downloaded. There are complete grammars available for several languages and there are a great number of tutorials and forums available for users and developers of the product.


[12] Java 1.4.2 language specification: http://java.sun.com/j2se/1.4.2/docs/api/


Appendix A: Java 1.4.2 Input/Output Classes

In the electronic version of this document all the keywords below are hyperlinked to the documentation for Java Version 1.4.2 on Sun’s web site. *Italicized* entries are interfaces; regular type entries are classes.

### Input Classes/Interfaces

<table>
<thead>
<tr>
<th>Class/Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>AudioFileReader</td>
</tr>
<tr>
<td>AudioInputStream</td>
</tr>
<tr>
<td>BufferedReader</td>
</tr>
<tr>
<td>ByteArrayInputStream</td>
</tr>
<tr>
<td>CipherInputStream</td>
</tr>
<tr>
<td>DataInputStream</td>
</tr>
<tr>
<td>DigestInputStream</td>
</tr>
<tr>
<td>FileCacheImageInputStream</td>
</tr>
<tr>
<td>FileInputStream</td>
</tr>
<tr>
<td>FileReader</td>
</tr>
<tr>
<td>FilterInputStream</td>
</tr>
<tr>
<td>FilterReader</td>
</tr>
<tr>
<td>ImageInputStream</td>
</tr>
<tr>
<td>ImageInputStreamImpl</td>
</tr>
<tr>
<td>ImageInputStreamSpi</td>
</tr>
<tr>
<td>ImageReader</td>
</tr>
<tr>
<td>ImageReaderSpi</td>
</tr>
<tr>
<td>ImageReaderWriterSpi</td>
</tr>
<tr>
<td>ImageReadParam</td>
</tr>
<tr>
<td>InputStream</td>
</tr>
<tr>
<td>InputStreamReader</td>
</tr>
<tr>
<td>JarInputStream</td>
</tr>
<tr>
<td>MemoryCacheImageInputStream</td>
</tr>
<tr>
<td>MidiFileReader</td>
</tr>
<tr>
<td>ObjectInput</td>
</tr>
<tr>
<td>ObjectInputStream</td>
</tr>
<tr>
<td>ObjectInputStream.GetField</td>
</tr>
<tr>
<td>ObjectInputValidation</td>
</tr>
<tr>
<td>PipedInputStream</td>
</tr>
<tr>
<td>PipedReader</td>
</tr>
</tbody>
</table>

### Output Classes/Interfaces

<table>
<thead>
<tr>
<th>Class/Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>AudioFileWriter</td>
</tr>
<tr>
<td>BufferedOutputStream</td>
</tr>
<tr>
<td>BufferedWriter</td>
</tr>
<tr>
<td>ByteArrayOutputStream</td>
</tr>
<tr>
<td>CipherOutputStream</td>
</tr>
<tr>
<td>DataOutputStream</td>
</tr>
<tr>
<td>DigestOutputStream</td>
</tr>
<tr>
<td>FileCacheImageOutputStream</td>
</tr>
<tr>
<td>FileOutputStream</td>
</tr>
<tr>
<td>FileWriter</td>
</tr>
<tr>
<td>FilterOutputStream</td>
</tr>
<tr>
<td>FilterWriter</td>
</tr>
<tr>
<td>ImageOutputStream</td>
</tr>
<tr>
<td>ImageOutputStreamImpl</td>
</tr>
<tr>
<td>ImageOutputStreamSpi</td>
</tr>
<tr>
<td>ImageWriteParam</td>
</tr>
<tr>
<td>ImageWriter</td>
</tr>
<tr>
<td>ImageWriterSpi</td>
</tr>
<tr>
<td>JarOutputStream</td>
</tr>
<tr>
<td>MemoryCacheImageOutputStream</td>
</tr>
<tr>
<td>MidiFileWriter</td>
</tr>
<tr>
<td>ObjectOutput</td>
</tr>
<tr>
<td>ObjectOutputStream</td>
</tr>
<tr>
<td>ObjectOutputStream.PutField</td>
</tr>
<tr>
<td>OutputStream</td>
</tr>
<tr>
<td>OutputStreamWriter</td>
</tr>
</tbody>
</table>

*Italicized* entries are interfaces; regular type entries are classes.
Appendix A (continued)

PipedOutputStream
PipedWriter
PrintStream
PrintWriter
StringWriter
Writer
ZipOutputStream

Other I/O-related Classes/Interfaces

EOFException
File
FileVersion
FileChannel
FileChannel.MapMode
FileChooserUI
FileDescriptor
FileDialog
FileFilter
FileHandler
FileFilter
FileLock
FileLock InterruptionException
FilenameFilter
FileNameMap
FileNotFoundException
FilePermission
FileSystemView
FileView
FileReader
FileChannel
FileChannel.MapMode
FileChannel.MapMode
FileDescriptor
FileDialog
FileFilter
FileHandler
FileFilter
FileLock
FileLock InterruptionException
FilenameFilter
FileNameMap
FileNotFoundException
FilePermission
FileSystemView
FileView
IIOByteBuffer
IIOException
IIOImage
IIOInvalidTreeException
IOMetadata
IOMetadataController
IOMetadataFormat
IOMetadataFormatImpl
IOMetadataNode
IOParam
IOParamController
IOREgistry
IOServiceProvider
IIOReadProgressListener
IIOReadUpdateListener
IIOReadWarningListener
IIOWriteProgressListener
IIOWriteWarningListener
ImageIO
InputMap
InputMapUIResource
InputContext
InputEvent
InputMethod
InputMethodContext
InputMethodDescriptor
InputMethodEvent
InputMethodHighlight
InputMethodListener
InputMethodRequests
InputSource
InputSubset
InputVerifier
IOException
JarFile
ObjectStreamClass
ObjectStreamConstants
ObjectStreamException
ObjectStreamField
ReadOnlyBufferException
StreamCorruptedException
StreamHandler
StreamPrintService
StreamPrintServiceFactory
StreamResult
StreamSource
StreamTokenizer
StringBuffer
StringTokenizer
WritableByteBuffer
WritableRaster
WritableByteArrayOutputStream
ZipFile
Appendix B: List of Java 1.4.2 Output Methods

Below is a listing of all output methods associated with output classes in Java

1.4.2. Many methods are polymorphic, often associated with many or most Java output classes.

checkError  setOutput
connect   write
createOutputStreamInstance  writeBit
createWriterInstance  writeBytes
flush   writeBoolean
flushBefore   writeByte
flushBits   writeChars
newLine   writeChar
prepareInsertEmpty  writeChars
prepareReplacePixels  writeClassDescriptor
prepareWriteEmpty  writeDouble
prepareWriteSequence  writeDoubles
print   writeFields
println   writeFloat
processImageProgress   writeFloats
processImageStarted   writeInt
processThumbnailComplete   writeInsert
processThumbnailProgress   writeLong
processThumbnailStarted   writeLongs
put   writeObjectOverride
putField   writeShort
putNextEntry   writeShorts
replaceImageMetadata   writeStreamHeader
replaceObject   writeToSequence
replacePixels   writeTo
replacePixels   writeUnshared
replaceStreamMetadata   writeUTF
reset
seek
Appendix C: Java 1.4.2 Output Classes and Associated Output Methods

AudioFileWriter()
write(AudioInputStream stream, AudioFileFormat.Type fileType, File out)
write(AudioInputStream stream, AudioFileFormat.Type fileType, OutputStream out)

BufferedOutputStream(OutputStream out)
BufferedOutputStream(OutputStream out, int size)
flush()
write(byte[] b, int off, int len)
write(int b)

BufferedWriter(Writer out)
BufferedWriter(Writer out, int sz)
flush()
newLine()
write(char[] cbuf, int off, int len)
write(int c)
write(String s, int off, int len)

ByteArrayOutputStream()
ByteArrayOutputStream(int size)
reset()
write(byte[] b, int off, int len)
write(int b)
writeTo(OutputStream out)

CharArrayWriter()
CharArrayWriter(int initialSize)
flush()
reset()
write(char[] c, int off, int len)
write(int c)
writeTo(Writer out)

CipherOutputStream(OutputStream os)
CipherOutputStream(OutputStream os, Cipher c)
flush()
write(byte[] b)
write(byte[] b, int off, int len)
write(int b)

DataOutputStream(OutputStream out)
flush()
write(byte[] b, int off, int len)
write(int b)
Appendix C (continued)

writeBoolean(boolean v)
writeByte(int v)
writeBytes(String s)
writeChar(int v)
writeChars(String s)
writeDouble(double v)
writeFloat(float v)
writeInt(int v)
writeLong(long v)
writeShort(int v)
writeUTF(String str)

DigestOutputStream(OutputStream stream, MessageDigest digest)
write(byte[] b, int off, int len)
write(int b)

FileCacheImageOutputStream(OutputStream stream, File cacheDir)
flushBefore(long pos)
seek(long pos)
write(byte[] b, int off, int len)
write(int b)

FileImageOutputStream(File f)
FileImageOutputStream(RandomAccessFile raf)
seek(long pos)
write(byte[] b, int off, int len)
write(int b)

FileOutputStream(File file)
FileOutputStream(File file, boolean append)
FileOutputStream(FileDescriptor fdObj)

FileOutputStream(String name)
FileOutputStream(String name, boolean append)
write(byte[] b)
write(byte[] b, int off, int len)
write(int b)

FileWriter(File file)
FileWriter(File file, boolean append)
FileWriter(FileDescriptor fd)
FileWriter(String fileName)
FileWriter(String fileName, boolean append)
Appendix C (continued)

FilterOutputStream(OutputStream out)
flush()
write(byte[] b)
write(byte[] b, int off, int len)
write(int b)

FilterWriter(Writer out)
flush()
write(char[] cbuf, int off, int len)
write(int c)
write(String str, int off, int len)

ImageOutputStreamImpl()
flushBits()
write(byte[] b)
write(byte[] b, int off, int len)
write(int b)
writeBit(int bit)
writeBits(long bits, int numBits)
writeBoolean(boolean v)
writeByte(int v)
writeBytes(String s)
writeChar(int v)
writeChars(char[] c, int off, int len)
writeChars(String s)
writeDouble(double v)
writeDoubles(double[] d, int off, int len)
writeFloat(float v)
writeFloats(float[] f, int off, int len)
writeInt(int v)
writeInts(int[] i, int off, int len)
writeLong(long v)
writeLongs(long[] l, int off, int len)
writeShort(int v)
writeShorts(short[] s, int off, int len)
writeUTF(String s)

ImageOutputStreamSpi()
ImageOutputStreamSpi(String vendorName, String version, Class outputClass)
createOutputStreamInstance(Object output)
createOutputStreamInstance(Object output, boolean useCache, File cacheDir)

ImageWriter(ImageWriterSpi originatingProvider)
Appendix C (continued)

prepareInsertEmpty(int imageIndex, ImageTypeSpecifier imageType, int width, int height, IIOMetadata imageMetadata, List thumbnails, ImageWriteParam param)
prepareReplacePixels(int imageIndex, Rectangle region)
prepareWriteEmpty(IIOMetadata streamMetadata, ImageTypeSpecifier imageType, int width, int height, IIOMetadata imageMetadata, List thumbnails, ImageWriteParam param)
prepareWriteSequence(IIOMetadata streamMetadata)
processImageProgress(float percentageDone)
processImageStarted(int imageIndex)
processThumbnailComplete()
processThumbnailProgress(float percentageDone)
processThumbnailStarted(int imageIndex, int thumbnailIndex)
replaceImageMetadata(int imageIndex, IIOMetadata imageMetadata)
replacePixels(Raster raster, ImageWriteParam param)
replacePixels(RenderedImage image, ImageWriteParam param)
replaceStreamMetadata(IIOMetadata streamMetadata)
setOutput(Object output)
write(IIOImage image)
write(IIOMetadata streamMetadata, IIOLImage image, ImageWriteParam param)
write(RenderedImage image)
writeInsert(int imageIndex, IIOLImage image, ImageWriteParam param)
writeToSequence(IIOLImage image, ImageWriteParam param)

ImageWriterSpi()
ImageWriterSpi(String vendorName, String version, String[] names, String[] suffixes, String[] MIMETypes, String writerClassName, Class[] outputTypes, String[] readerSpiNames, boolean supportsStandardStreamMetadataFormat, String nativeStreamMetadataFormatName, String nativeStreamMetadataFormatClassName, String[] extraStreamMetadataFormatNames, String[] extraStreamMetadataFormatClassNames, boolean supportsStandardImageMetadataFormat, String nativeImageMetadataFormatName, String nativeImageMetadataFormatClassName, String[] extraImageMetadataFormatNames, String[] extraImageMetadataFormatClassNames)
createWriterInstance()
createWriterInstance(Object extension)

JarOutputStream(OutputStream out)
JarOutputStream(OutputStream out, Manifest man)
putNextEntry(ZipEntry ze)

MemoryCacheImageOutputStream(OutputStream stream)
flushBefore(long pos)
write(byte[] b, int off, int len)
write(int b)
Appendix C (continued)

MidiFileWriter()
write(Sequence in, int fileType, File out)
write(Sequence in, int fileType, OutputStream out)

ObjectOutputStream()
ObjectOutputStream(OutputStream out)
flush(
putFields()
replaceObject(Object obj)
write(byte[] buf)
write(byte[] buf, int off, int len)
write(int val)
writeBoolean(boolean val)
writeByte(int val)
writeBytes(String str)
writeChar(int val)
writeChars(String str)
writeClassDescriptor(ObjectStreamClass desc)
writeDouble(double val)
writeFields()
writeFloat(float val)
writeInt(int val)
writeLong(long val)
writeObject(Object obj)
writeObjectOverride(Object obj)
writeShort(int val)
writeStreamHeader()
writeUnshared(Object obj)
writeUTF(String str)

ObjectOutputStream.PutField()
put(String name, boolean val)
put(String name, byte val)
put(String name, char val)
put(String name, double val)
put(String name, float val)
put(String name, int val)
put(String name, long val)
put(String name, Object val)
put(String name, short val)
write(ObjectOutput out)

OutputStream()
flush()
Appendix C (continued)

write(byte[] b)
write(byte[], int off, int len)
write(int b)

OutputStreamWriter(OutputStream out)
OutputStreamWriter(OutputStream out, Charset cs)
OutputStreamWriter(OutputStream out, CharsetEncoder enc)
OutputStreamWriter(OutputStream out, String charsetName)
flush()
write(char[] cbuf, int off, int len)
write(int c)
write(String str, int off, int len)

PipedOutputStream()
PipedOutputStream(PipedInputStream snk)
connect(PipedInputStream snk)
flush()
write(byte[], int off, int len)
write(int b)

PipedWriter()
PipedWriter(PipedReader snk)
connect(PipedReader snk)
flush()
write(char[] cbuf, int off, int len)
write(int c)

PrintStream(OutputStream out)
PrintStream(OutputStream out, boolean autoFlush)
PrintStream(OutputStream out, boolean autoFlush, String encoding)
checkError()
flush()
print(boolean b)
print(char c)
print(char[] s)
print(double d)
print(float f)
print(int i)
print(long l)
print(Object obj)
print(String s)
println()
println(boolean x)
println(char x)
Appendix C (continued)

println(char[] x)
println(double x)
println(float x)
println(int x)
println(long x)
println(Object x)
println(String x)
write(byte[] buf, int off, int len)
write(int b)

PrintWriter(OutputStream out)
PrintWriter(OutputStream out, boolean autoFlush)
PrintWriter(Writer out)
PrintWriter(Writer out, boolean autoFlush)
checkError()
flush()
print(boolean b)
print(char c)
print(char[] s)
print(double d)
print(float f)
print(int i)
print(long l)
print(Object obj)
print(String s)
println()
println(boolean x)
println(char x)
println(char[] x)
println(double x)
println(float x)
println(int x)
println(long x)
println(Object x)
println(String x)
write(char[] buf)
write(char[] buf, int off, int len)
write(int c)
write(String s)
write(String s, int off, int len)

StringWriter()
StringWriter(int initialSize)
flush()
Appendix C (continued)

write(char[] buf, int off, int len)
write(int c)
write(String s)
write(String s, int off, int len)

Writer()
Writer(Object lock)
flush()
write(char[] buf)
write(char[] buf, int off, int len)
write(int c)
write(String s)
write(String s, int off, int len)

ZipOutputStream(OutputStream out)
putNextEntry(ZipEntry e)
write(byte[] b, int off, int len)
### Appendix D: List of Java 1.4.2 tokens

<table>
<thead>
<tr>
<th></th>
<th>Token</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ABSTRACT</td>
</tr>
<tr>
<td>2</td>
<td>ARRAY_DECLARATOR</td>
</tr>
<tr>
<td>3</td>
<td>ARRAY_INIT</td>
</tr>
<tr>
<td>4</td>
<td>ASSIGN</td>
</tr>
<tr>
<td>5</td>
<td>BAND</td>
</tr>
<tr>
<td>6</td>
<td>BAND_ASSIGN</td>
</tr>
<tr>
<td>7</td>
<td>BLOCK</td>
</tr>
<tr>
<td>8</td>
<td>BNOT</td>
</tr>
<tr>
<td>9</td>
<td>BOR</td>
</tr>
<tr>
<td>10</td>
<td>BOR_ASSIGN</td>
</tr>
<tr>
<td>11</td>
<td>BSR</td>
</tr>
<tr>
<td>12</td>
<td>BSR_ASSIGN</td>
</tr>
<tr>
<td>13</td>
<td>BXOR</td>
</tr>
<tr>
<td>14</td>
<td>BXOR_ASSIGN</td>
</tr>
<tr>
<td>15</td>
<td>CASE_GROUP</td>
</tr>
<tr>
<td>16</td>
<td>CHAR_LITERAL</td>
</tr>
<tr>
<td>17</td>
<td>CLASS_DEF</td>
</tr>
<tr>
<td>18</td>
<td>COLON</td>
</tr>
<tr>
<td>19</td>
<td>COMMA</td>
</tr>
<tr>
<td>20</td>
<td>CTOR_CALL</td>
</tr>
<tr>
<td>21</td>
<td>CTOR_DEF</td>
</tr>
<tr>
<td>22</td>
<td>DEC</td>
</tr>
<tr>
<td>23</td>
<td>DIV</td>
</tr>
<tr>
<td>24</td>
<td>DIV_ASSIGN</td>
</tr>
<tr>
<td>25</td>
<td>DOT</td>
</tr>
<tr>
<td>26</td>
<td>ELIST</td>
</tr>
<tr>
<td>27</td>
<td>EMPTY_STAT</td>
</tr>
<tr>
<td>28</td>
<td>EQUAL</td>
</tr>
<tr>
<td>29</td>
<td>ESC</td>
</tr>
<tr>
<td>30</td>
<td>EXPONENT</td>
</tr>
<tr>
<td>31</td>
<td>EXPR</td>
</tr>
<tr>
<td>32</td>
<td>EXTENDS_CLAUSE</td>
</tr>
<tr>
<td>33</td>
<td>FINAL</td>
</tr>
<tr>
<td>34</td>
<td>FLOAT_SUFFIX</td>
</tr>
<tr>
<td>35</td>
<td>FOR_CONDITION</td>
</tr>
<tr>
<td>36</td>
<td>FOR_INIT</td>
</tr>
<tr>
<td>37</td>
<td>FOR_ITERATOR</td>
</tr>
<tr>
<td>38</td>
<td>GE</td>
</tr>
<tr>
<td>39</td>
<td>GT</td>
</tr>
<tr>
<td>40</td>
<td>HEX_DIGIT</td>
</tr>
<tr>
<td>41</td>
<td>IDENT</td>
</tr>
<tr>
<td>42</td>
<td>IMPLEMENTS_CLAUSE</td>
</tr>
<tr>
<td>43</td>
<td>IMPORT</td>
</tr>
<tr>
<td>44</td>
<td>INC</td>
</tr>
<tr>
<td>45</td>
<td>INDEX_OP</td>
</tr>
<tr>
<td>46</td>
<td>INSTANCE_INIT</td>
</tr>
<tr>
<td>47</td>
<td>INTERFACE_DEF</td>
</tr>
<tr>
<td>48</td>
<td>LABELED_STAT</td>
</tr>
<tr>
<td>49</td>
<td>LAND</td>
</tr>
<tr>
<td>50</td>
<td>LBRACK</td>
</tr>
<tr>
<td>51</td>
<td>LCURLY</td>
</tr>
<tr>
<td>52</td>
<td>LE</td>
</tr>
<tr>
<td>53</td>
<td>LNOT</td>
</tr>
<tr>
<td>54</td>
<td>LOR</td>
</tr>
<tr>
<td>55</td>
<td>LPAREN</td>
</tr>
<tr>
<td>56</td>
<td>LT</td>
</tr>
<tr>
<td>57</td>
<td>METHOD_CALL</td>
</tr>
<tr>
<td>58</td>
<td>METHOD_DEF</td>
</tr>
<tr>
<td>59</td>
<td>MINUS</td>
</tr>
<tr>
<td>60</td>
<td>MINUS_ASSIGN</td>
</tr>
<tr>
<td>61</td>
<td>ML_COMMENT</td>
</tr>
<tr>
<td>62</td>
<td>MOD</td>
</tr>
<tr>
<td>63</td>
<td>MOD_ASSIGN</td>
</tr>
<tr>
<td>64</td>
<td>MODIFIERS</td>
</tr>
<tr>
<td>65</td>
<td>NOT_EQUAL</td>
</tr>
<tr>
<td>66</td>
<td>NUM_DOUBLE</td>
</tr>
<tr>
<td>67</td>
<td>NUM_FLOAT</td>
</tr>
<tr>
<td>68</td>
<td>NUM_INT</td>
</tr>
<tr>
<td>69</td>
<td>NUM_LONG</td>
</tr>
<tr>
<td>70</td>
<td>OBJBLOCK</td>
</tr>
<tr>
<td>71</td>
<td>PACKAGE_DEF</td>
</tr>
<tr>
<td>72</td>
<td>PARAMETER_DEF</td>
</tr>
<tr>
<td>73</td>
<td>PARAMETERS</td>
</tr>
<tr>
<td>74</td>
<td>PLUS</td>
</tr>
<tr>
<td>75</td>
<td>PLUS_ASSIGN</td>
</tr>
<tr>
<td>76</td>
<td>POST_DEC</td>
</tr>
<tr>
<td>77</td>
<td>POST_INC</td>
</tr>
<tr>
<td>78</td>
<td>QUESTION</td>
</tr>
<tr>
<td>79</td>
<td>RBRACK</td>
</tr>
<tr>
<td>80</td>
<td>RCU LRY</td>
</tr>
<tr>
<td>81</td>
<td>RPAREN</td>
</tr>
<tr>
<td>82</td>
<td>SEMI</td>
</tr>
<tr>
<td>83</td>
<td>SL</td>
</tr>
<tr>
<td>84</td>
<td>SL_ASSIGN</td>
</tr>
</tbody>
</table>
Appendix D (continued)

85. SL_COMMENT
86. SLIST
87. SR
88. SR_ASSIGN
89. STAR
90. STAR_ASSIGN
91. STATIC_INIT
92. STRICTFP
93. STRING_LITERAL
94. SUPERCTOR_CALL
95. TYPE
96. TYPECAST
97. UNARY_MINUS
98. UNARY_PLUS
99. VARIABLE_DEF
100. VOCAB
101. WS
102. abstract
103. assert
104. boolean
105. break
106. byte
107. case
108. catch
109. char
110. class
111. continue
112. default
113. do
114. double
115. else
116. extends
117. false
118. final
119. finally
120. float
121. for
122. if
123. implements
124. import
125. instanceof
126. int
127. interface
128. long
129. native
130. new,
131. null
132. package
133. private
134. protected
135. public
136. return
137. short
138. static
139. strictfp
140. super
141. switch
142. synchronized
143. this
144. threadsafe
145. throw
146. throws
147. transient
148. true
149. try
150. void
151. volatile
152. while
153. byvalue
154. cast
155. const
156. future
157. generic
158. goto
159. inner
160. operator
161. outer
162. rest
163. var
Appendix E: *for*, *while* and *do* structures tested with IOScan

1. `vartype x;`  
   `... for (x = n; x bop n; iter) { ... }`

2. `for (vartype x = n; x bop n; iter) { ... }`

3. `vartype x;`  
   `... x = n;`  
   `... while (x bop n) { ... iter ... }`

4. `vartype x = n;`  
   `... while (x bop n) { ... iter ... }`

5. `vartype x;`  
   `... x = n;`  
   `... do { ... iter ... } while (x bop n)`

6. `vartype x = n;`  
   `... do { ... iter ... } while (x bop n)`

Explanation:

*iter* can take any of the following forms:

- `x++` | `x--` | `++x` | `--x` | `x op= n` | `x = x op n`

*op* are standard arithmetic operators:

- `+`, `-`, `*`, `/`

*bop* are standard Boolean operators:

- `<`, `>`, `=`, `<=`, `>=`, `!=`

*vartype* can be any legal Java primitive numeric type

*n* can be any permissible numeric value, depending on vartype

... represents additional in-line code
import java.io.*;
import java.util.Hashtable;
import antlr.collections.AST;
import antlr.collections.ASTEnumeration;
import antlr.collections.impl.*;
import antlr.debug.misc.*;
import ant1r.*;
import java.awt.event.*;
import java.lang.Math;

public class IOScan {
    //Class-level declaration
    static boolean showTree = false;  //show tree in frame or not
    static Hashtable outHT = new Hashtable();    //hash Java output methods
    static Hashtable blockHT = new Hashtable();  //hash: while, for, do.
    static Hashtable tokensHT = new Hashtable(); //hash: all AST tokens
    static AST fullTree = null; //full AST tree

    //main method - dispatches methods for building hash tables, main AST
    // tree and locating iterated Java output methods
    public static void main(String[] args) throws Exception {
        //declarations/initialization
        AST t = null;
        Vector roots = new Vector(10);
        ASTEnumeration e = null;

        //build hash tables
        hashOutputMethods();
        hashJavaLoops();
        hashTokens();

        //build AST tree and walk it, looking for loops
        t = buildAST(args);
        fullTree = t;
        if (t != null) {
            e = findLoops(roots, t);  //builds an enum of AST
            if (e != null) {

                System.out.println("main complete.");
            }
        }
    }

    //recursive method - depth-first search through tree looking for loops
    public static ASTEnumeration findLoops(Vector v, AST t) {
        //declarations and initialization
        AST thisNode, child;
        thisNode = t;

        while (thisNode != null) {
            child = thisNode.getFirstChild();
            if (child != null) {
                findLoops(v, child); //recursion occurs here
            }
            v.addElement(thisNode);
            checkLoops(thisNode, t); //if thisNode starts loop, look for I/O
            thisNode = thisNode.getNextSibling();
        }
    }
}

Copyright © 2005, Open-source GPL license, Bill Madden (bmadden@bergen.edu)
Appendix F (continued)

```java
60     } //while
61     return new ASTEnumerator(v);
62 }
63
64 } //findLoops
65
66 //check to see if thisNode starts a loop (for, while, do)
67 public static void checkLoops(AST node, AST t) {
68     //declarations and initialization
69     AST child = null, sibling = null, varType = null;
70     AST ident = null, assign = null, expr = null;
71     AST subtree = null;
72     ASTEnumeration st = null;
73     boolean foundVarDef = false;
74     String nodeText = node.getText();
75
76     //if node is start of a loop structure, look for output methods
77     if (blockHT.contains(nodeText)) {
78         System.out.println("Found: " + nodeText);
79     }
80     //check subtree for output methods
81     subtree = node.getFirstChild(); //must go inside of structure
82     st = findOutputNodes(subtree);
83     //if no output methods found, move on
84     if (!st.hasMoreNodes()) {
85         System.out.println(" No output nodes found");
86     } else { //show output nodes, branch to process for, while & do
87         while (st.hasMoreNodes()) {
88             System.out.println(" Output node in loop: " +
89                 st.nextNode());
90         }
91     //if output method found, branch: process for-while-do loops
92     if (nodeText.equals("for")) {
93         processFor(node);
94     } else if (nodeText.equals("while")) {
95         // processWhileDo(node, t); //logic very similar for both
96     } else {
97         processWhileDo(node, t);
98         //System.out.println("Error: for-while-do ONLY coded");
99     } //if there are output nodes inside this current loop
100 } //if the current node is a loop structure
101
102 } //checkLoops
103
104 //Process for loops: captures init, exit and iter conditions that define
105 // the execution of the for loop
106 public static void processFor(AST loopNode) {
107     //presently, we assume that for_loop init, exit (cond) and iterator
108     // have no dependencies.
109     //declarations and initialization
110     AST init, cond, iter;
111     AST elist, expr;
112     AST initoper, condOper, iterOper;
113     AST initIdent, initVal, tmp, conIdent, condVal, iterIdent,
114     arithOper, iterVal;
115
116     arithOper = null;
```
iterVal = null;
String strCond = "";  //represents exit condition operator

//parse nodes in FOR_INIT, FOR_CONDITION and FOR_ITERATOR subtrees
init = loopNode.getFirstChild();  //walking default structure of
cond = init.getNextSibling();  //for-loop with no dependencies and
iter = cond.getNextSibling();  //simple assignment and boolean
tmp = init.getFirstChild();  //expressions for init/cond/iter
if (tmp.getText().equals("ELIST")) {
  expr = tmp.getFirstChild();
  initOper = expr.getFirstChild();
  initIdent = initOper.getFirstChild();
  initVal = initIdent.getNextSibling();
} else {  //VAR_DEF is occurring inside for definition
  expr = tmp.getFirstChild();  //Modifiers
  tmp = tmp.getNextSibling();  //Type
  initIdent = tmp.getNextSibling();  //Ident name
  initOper = initIdent.getNextSibling();  //init operator
  expr = expr.getFirstChild();  //EXPR node
  initVal = initIdent.getFirstChild();  //init val of loop-control ident
}
expr = cond.getFirstChild();  //keep walking to locate
condOper = expr.getFirstChild();  //iter node
condIdent = condOper.getFirstChild();
condVal = condIdent.getNextSibling();
elist = iter.getFirstChild();
expr = elist.getFirstChild();
iterOper = expr.getFirstChild();
iterIdent = iterOper.getFirstChild();

//declarations for determining loop direction and behavior
int itDir = 0;  //+int moves right (on number line); -int moves left
double dblInit, dblCond, dblIter;  //converts AST nodes to numerics
dblInit = dblCond = dblIter = 0.0;

//convert value nodes to numerics - doubles are least restrictive
//next 7 lines convert the cond (exit_condition) node to a double
if (condVal.getText().equals("-")) {  //in case of unary minus
  tmp = condVal.getFirstChild();
  dblCond = -(Double.parseDouble(tmp.getText()));
} else if (condVal.getText().equals("+")) {  //rare case: unary plus
  tmp = condVal.getFirstChild();
  dblCond = Double.parseDouble(tmp.getText());
} else dblCond = Double.parseDouble(condVal.getText());

//next 7 lines convert init_condition to a double
if (initVal.getText().equals("-")) {  //in case of unary minus
  tmp = initVal.getFirstChild();
  dblInit = -(Double.parseDouble(tmp.getText()));
} else if (initVal.getText().equals("+")) {  //rare case: unary plus
  tmp = initVal.getFirstChild();
  dblInit = Double.parseDouble(tmp.getText());
} else dblInit = Double.parseDouble(initVal.getText());

//evaluate iter expression, result will be a double
String strOper = iterOper.getText();  //if iterOper=null report Error
if (strOper.equals("=")) {  //assumes expression form: x = x op n
  iterVal = iterOper.getFirstChild();
  x op n
Appendix F (continued)

    arithOper = iterVal.getNextSibling(); //this can be: +, -, *, /
    iterVal = arithOper.getFirstChild();
    iterVal = iterVal.getNextSibling();
    dblIter = Double.parseDouble(iterVal.getText());
} else {
    dblIter = getIter(iterOper);
} //if

if (arithOper != null) {
    strOper = arithOper.getText() + "=
    }

//There are 9 possible outcomes for a direction of the iterator:
//see comments in iterDirect method signature (line 470 approx)
itDir = iterDirect(strOper, dblIter, dblInit);
strCond = condOper.getText();

//We now have all the values needed for analysis of for loops:
// dbllnit = initial starting value of loop control identifier
// dblliter and strOper = iterator value and operation
// dblCond = exit value of loop control identifier
//Logic is similar to that in the while/do loop analysis.
//display feedback analysis (line 530 approx)
feedback(dblInit, dblIter, dblCond, itDir, strCond);

} //processFor

//Process while and do loops
public static void processWhileDo(AST loopNode, AST t) {

    //presently we assume a while or do loop employing a loop control
    //variable that has no dependencies on its exit_condition.

    //declarations and initialization
    int condType;
    //ASTEnumeration e;
    AST init, cond, iter;
    AST sList, expr;
    AST initOper, condOper, iterOper, arithOper;
    AST initIdent, initVal, tmp, condIdent, condVal, iterIdent, iterVal;
    iterVal = null;
    arithOper = null;
    String strCond; //represents of exit condition operator

    //the following will be used to convert AST nodes to numerics so we
    // can evaluate loop direction and behavior
    int itDir = 0;
    double dblInit, dblCond, dblIter;
    dblInit =dblCond = dblIter = 0.0;

    //There are two major nodes inside while/do loops: EXPR and SLIST
    // EXPR contains the exit condition for the loop
    // SLIST contains the body of the loop, including the iterator
    if (loopNode.getText().equals("while")) {
        expr = loopNode.getFirstChild() //while loop
        sList = expr.getNextSibling();
    } else {
        sList = loopNode.getFirstChild(); //order of two major nodes is
Appendix F (continued)

```java
expr = sList.getNextSibling(); // reversed in do-loop
}

// walk the EXPR node: parenthetical exit condition following while/do
// e = evalTree(expr); - eventually we will have a general EXPression evaluator
condOper = expr.getFirstChild();
condIdent = condOper.getFirstChild();
condVal = condIdent.getNextSibling();

// Convert cond (exit_condition) to double, testing for unary
// minus and unary plus along the way
if (condVal.getText().equals("-")) {
    // in case of unary minus
    tmp = condVal.getFirstChild();
    dblCond = -(Double.parseDouble(tmp.getText()));
} else if (condVal.getText().equals("+")) {
    // rare case: unary plus
    tmp = condVal.getFirstChild();
    dblCond = Double.parseDouble(tmp.getText());
} else dblCond = Double.parseDouble(condVal.getText());

condType = condVal.getType(); // 138=int/141=float/142=long/
// 143=double, but compiler does
// implicit casting when it sets actual
// coded val into loopCtrlldent

if (exitVal.getText().equals("EXPR")) then the expression is more
// complex and we will need a more general EXPression evaluator
// Now we have the loop control variable (AST loopCtrlldent),
// its exit value (AST exitVal) and type (int loopCtrlType) - though
// this last is implicitly cast by compiler into loopCtrlldent's
// actual declared type.

// Next we search SLIST locating the EXPR that contains the iterator.
// Look for three forms of iterator: POST_INC, ASSIGN and PLUS_ASSIGN.
// Convert iter node to double
AST sListChild = sList.getFirstChild(); // drop into SLIST first
iterOper = findIterOper(sListChild, condIdent);
String strOper = iterOper.getText();
if (strOper.equals("=")) { // assumes expression form: x = x op n
    iterVal = iterOper.getFirstChild();
    arithOper = iterVal.getNextSibling(); // this could be: +, -, *, /
    iterVal = arithOper.getFirstChild();
    iterVal = iterVal.getNextSibling();
    // check for a unary minus
    if (iterVal.getText().equals("-")) {
        tmp = iterVal.getFirstChild();
        dblIter = -(Double.parseDouble(tmp.getText()));
    } else if (iterVal.getText().equals("+")) { // rare: unary plus
        tmp = iterVal.getFirstChild();
        dblIter = Double.parseDouble(tmp.getText());
    } else dblIter = Double.parseDouble(iterVal.getText());
} else {
    dblIter = getIter(iterOper);
} // if

if (arithOper !=null) {
    strOper = arithOper.getText() + "=";
} // if

// Next we locate the VARIABLE_DEF and/or EXPR that defines init
// condition. findInitVal locates only declarations/expressions of
```
Appendix F (continued)

```java
//the form: int x = n;
//int x; x = n;
//int x; x += n; (or *=, /=, -=)
initVal = null;
initVal = findInitVal(fullTree, condIdent, loopNode, initVal);
if (initVal.getText().equals("-")) {
    tmp = initVal.getFirstChild();
    dblInit = -(Double.parseDouble(tmp.getText()));
} else if (initVal.getText().equals("+")) {//rare: unary plus
    tmp = initVal.getFirstChild();
    dblInit = Double.parseDouble(tmp.getText());
} else dblInit = Double.parseDouble(initVal.getText());

//There are 9 possible outcomes for a direction of the iterator:
//see comments in iterDirect method signature (line 470 approx.)
itDir = iterDirect(strOper, dblIter, dblInit);
strCond = condOper.getText();

//We now have all the values needed for analysis of while/do loops:
// dblInit = initial starting value of loop control identifier
// dblIter and strOper = iterator value and operation
// dblCond = exit value of loop control identifier
//Logic is similar to that in the for loop analysis.
//display feedback analysis (line 530 approx)
feedback(dblInit, dblIter, dblCond, itDir, strCond);

} //processWhileDo

//recursive method that starts walking entire AST tree finding first the
// VARIABLE_DEF of condIdent (the exit_condition of a while/do loop),
// examining it for an initial value of condIdent, then continuing to
// walk forward through AST t looking for any EXPRESSIONs involving a re-
// valuation of condIdent. It stops looking when it encounters loopNode
// (the current while/do loop that spawned this search).
public static AST findInitVal(AST t, AST condIdent, AST loopNode, AST result) {
    //Declarations and initialization
    AST thisNode, child; //needed for recursion
    String nodeText;
    thisNode = t;
    nodeText = thisNode.getText();
    if (nodeText.equals("VARIABLE_DEF")) {
        AST tmp;
        tmp = thisNode.getFirstChild(); //gets MODIFIERS node
        tmp = tmp.getNextSibling(); //gets TYPE node
        tmp = tmp.getNextSibling(); //gets IDENT (check against condIdent
        if (tmp.getText().equals(condIdent.getText())) {
            tmp = tmp.getFirstChild(); //gets initial initVal
            thisNode = thisNode.getNextSibling();
            nodeText = thisNode.getText();
        }
    }
}
```
// walk forward from VARIABLE_DEF looking for EXPRESSIONs that might
// re-evaluate the initVal identified in VARIABLE_DEF
if (nodeText.equals("EXPR")) {
    AST tmp; // subnodes
    String snText; // subnode text
    tmp = thisNode.getFirstChild();
    snText = tmp.getText();
    if (snText.equals("++") || snText.equals(" + =") || snText.equals("=")) {
        tmp = tmp.getFirstChild();
        if (tmp.getText().equals(condldent.getText())) {
            result = tmp.getNextSibling(); // get initVal
        } //if
        thisNode = thisNode.getNextSibling();
    } //if

    // recursive loop
    while (thisNode != null) {
        if (thisNode.equalsTree(loopNode)) { // if loopNode, end search
            return result;
        } else {
            child = thisNode.getFirstChild();
            if (child != null) { // recursive call is on next line
                result = findInitVal(child, condldent, loopNode,
                result);
            } //if
        } //if
        thisNode = thisNode.getNextSibling();
    } //if

    // initialize method that walks portion of tree finding an iterator node
    // within a while/do loop (POST_INC, ASSIGN or PLUS_ASSIGN); initially,
    // condldent is the iterator Identifier.
    public static AST findIterOper(AST sList, AST condldent) {
        // declarations and initialization
        AST thisNode, child, result;
        String nodeText;
        result = null;
        thisNode = sList;
        nodeText = thisNode.getText();

        // possible iterator expressions
        if (nodeText.equals("++") || nodeText.equals("+-") || nodeText.equals("=")) {
            child = thisNode.getFirstChild();
            if (child.getText().equals(condldent.getText())) {
                //...
Appendix F (continued)

```java
result = thisNode;
return result;
} //if
} //if

//recursive loop
while (thisNode != null) {
    child = thisNode.getFirstChild();
    if (child != null) {
        result = findIterOper(child, condIdent);  //recursive call
    } //if
    thisNode = thisNode.getNextSibling();
} //while

return result;

} //findIterOper

//For simple iterator expressions of the form:
//   x++, x--, ++x, --x, x+=n, x-=n, x*=n, x/=n
// this finds n. Expressions of the form x = x op n are processed in-
// line before this call. ++x and --x are structurally identical to x++
// and x-- in the AST tree and are processed in this code identically
public static double getIter(AST i) {

    //Declarations
    AST tmp = null;
    double result = 0.0;
    String s = i.getText();

    if (s.equals("++")) {
        //process x++
        result = 1.0;
    } else if (s.equals("+=") || s.equals("-=") || s.equals("*=") || s.equals("/=")) {
        //process x '+-*/=' n. Watch out: in odd cases this could
        //be an expression like: x *= -1;
        tmp = i.getFirstChild();
        tmp = tmp.getNextSibling();
        if (tmp.getText().equals("-"))  //in case of unary minus
            result = -(Double.parseDouble(tmp.getText())) ;
        else if (tmp.getText().equals("+")) {  //rare case: unary plus
            tmp = tmp.getFirstChild();
            result = Double.parseDouble(tmp.getText());  //iter found
        } else {
            //process x--
            result = -1.0;
        } //if
    }

    return result;
}

//defines 9 possible conditions plus one error condition for direction of
//iterator - see comments below
public static int iterDirect(String o, double n, double i) {
    int result = 0;

    //x++ or x += n or x-= -n; also covers x = x + n and x = x - (-n)
```
```java
if ((o.equals("++")))
|| (o.equals("+=") && n > 0)
|| (o.equals("-=") && n < 0))
result = 1; //iterator moves additively toward +inf
//with initVal > 0: x*=n (n>l) or x/=n (0<n<l);
// also covers x=x op n equivalents
else if (((o.equals("*=") && n > 1)
|| (o.equals("=/") && (n > 0 && n < 1))) && i > 0)
result = 1; //iterator moves multiplicatively toward +inf
//x-- or x+= -n or x-=n; also covers x=x+-(-n) and x=x-n
else if ((o.equals("--"))
|| (o.equals("+=") && n < 0) || (o.equals("-=") && n > 0))
result = 2; //iterator moves additively toward -inf
//with initVal < 0: x*=n (n>l) or x/=n (0<n<l);
// also covers x=x op n equivalents
else if (((o.equals("*=") && n > 1)
|| (o.equals("=/") && (n > 0 && n < 1))) && i < 0)
result = 2; //iterator moves multiplicatively to -inf
//with initVal = 0: x*=n or x/=n;
// also covers x=x op n equivalents - result is 0
else if ((o.equals("*=")
|| o.equals("/=")) && i == 0)
result = 7; //iterator never leaves 0
//x*=n (0<n<l) or x/=n (n>l); also covers x=x op n equivalents
else if (((o.equals("*=")) && (n > 0 && n < 1))
|| (o.equals("/=")) && n > 1))
result = 3; //iterator converges toward 0 from + or from -
//x*=n (-l<n<0) or x/=n (n<-l); also covers x=x op n equivalents
else if ((o.equals("*=")) && (n > -1 && n < 0))
|| (o.equals("/=")) && n < -1))
result = 4; //iterator converges toward 0 from both sides
//x*=n (n<-l) or x/=n (-1<n<0); also cover x=x op n equivalents
else if ((o.equals("*=")) && n < -1)
|| (o.equals("/=")) && (n > -1 && n < 0)))
result = 5; //iterator alternates growing toward -inf and +inf
//x*= -1 or x/= -1;
// also covers x=x op n equivalents - causes iteration from x to -x
else if ((o.equals("*=")
|| o.equals("/=")) && n == -1)
result = 6; //iterator oscillates: -initVal and +initVal (infinite)
//x+=0 or x-=0 or x+=1 or x-=1;
// also covers x=x op n equivalents - result is always x
else if (((o.equals("+=')) || o.equals("=-")) && n == 0)
|| (o.equals("*=")) || o.equals("/=")) && n == 1))
result = 7; //iterator is stuck at initVal (infinite)
//special case: x*=0 - becomes 0 in 1 iteration
else if (o.equals("*=")) && n == 0)
result = 8;
//special case: x/=0; iterator becomes undefined
else if (o.equals("/=")) && n == 0)
result = 9;
//undefined error
else result = -1;
return result;
```
Appendix F (continued)

```java
public static void feedback(double n, double t, double c, int d, String o) {
    // logic which decides if for loop is properly formed
    // check basic structure of loop

    // Warnings:
    // Use of '!=' operator in an loop exit condition
    if (o.equals("!=")) {
        System.out.println("Warning: using '!=' as the operator for the exit condition from a loop");
        System.out.println("can cause problems if the loop control variable doesn't land right on the");
        System.out.println("exit value. Usually, '<=' or '>=' are preferred.");
    } //if

    // Iterators involving *= or /= that converge toward 0
    if (d == 3) {
        System.out.println("This iterator converges toward 0 from either the positive direction or the");
        System.out.println("negative direction. Make sure the relation of the initial and exit");
        System.out.println("conditions of the loop are appropriate for this situation.");
    } else if (d == 4) {
        System.out.println("This iterator converges toward zero, alternating between negative and positive");
        System.out.println("values on each iteration. Make sure the relation of the initial and exit");
        System.out.println("conditions of the loop are appropriate for this situation, or choose a simpler");
        System.out.println("method of iterating from the start to the finish of the loop.");
    } //if

    // Properly formed loop
    if ((n < c && (o.equals("<") || o.equals("<="))) && d == 1)
        System.out.println("Basic structure of loop is okay");
    else if ((n > c && (o.equals(">") || o.equals(">=") && d == 2)) {
        System.out.println("Exit condition boolean operator is the inverse of what it should be");
        System.out.println("conditions of the loop are appropriate for caused loop termination without execution of loop body");
        System.out.println("Loop iterator headed in the wrong direction");
    } else if (n < c && d == 2 || n > c && d == 1) {
        System.out.println("Loop init and exit are same values!");
    } //if

    // Drop through and test if entrance and exit conditions are equal
    if (n == c) {
        System.out.println("Loop init and exit are same values!");
    } //if

    // Values of d below result from unusual iterator expressions and are most likely semantic errors in the source being examined
    if (d == 5) {
```

System.out.println(" Iterator alternates between negative and
positive values as it increases");
System.out.println(" in absolute value. Be sure the relation
between initial and exit conditions");
System.out.println(" is appropriate for this situation, or
choose a simpler way to iterate from the");
System.out.println(" start to the finish of the loop.");
 } else if (d == 6) {
 System.out.println(" Iterator alternates between same neg/pos
values - infinite loop");
 } else if (d == 7) {
 System.out.println(" Iterator is not changing - infinite loop");
 } else if (d == 8) {
 System.out.println(" Iterator is of the form 'x *= 0' which goes
to 0 after one iteration");
 } else if (d == 9) {
 System.out.println(" Iterator is of the form 'x /= 0' which is
undefined");
 } //if

//Drop through and test to see if there is excessive looping,
//at least for +/- loops; arbitrarily set at >1000000 loops
if ((d == 1 || d == 2) & Math.abs(c - n) > Math.abs(t) * 1000000) {
 System.out.println(" This output will execute more than 1 million
times!");
}

//Other loop tests can be coded here

} //feedback

//builds enumeration from walk through AST
public static ASTEnumeration findOutputNodes(AST t) {
 Vector roots = new Vector(10);
 if (t == null) return null;
 searchSubtrees(roots, t);
 return new ASTEnumerator(roots);
} //findOutputNodes

//recursive method called by findOutputNodes()
private static boolean searchSubtrees(Vector v, AST t) {
 AST thisNode, child;
 //AST varDefChild, varDefNextSibling, identSibling;
 boolean itemFound = false;
 thisNode = t;

 //recursive loop
 while (thisNode != null) {
 child = thisNode.getFirstChild();
 if (child != null) {
 itemFound = searchSubtrees(v, child); //recursive call
 } //if
 if (outHT.contains(thisNode.getText())) {
 itemFound = true;
 v.addElement(thisNode);
 } //if
 thisNode = thisNode.getNextSibling();
 } //while

return itemFound;
Appendix F (continued)

```java
public static AST buildAST(String[] args) throws Exception {
    String f;
    AST t = null;  // AST is an antlr class that defines structure of an
                   // Abstract Syntax Tree

    f = "c:/antlr/SimpleLoop.java";  //<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    t = parseFile(f, new BufferedReader(new FileReader(f)));  //<<<<<<<

    try {
        if (args.length > 0) {  //if we have at least 1 command-line
            System.err.println("Parsing... ");
            for(int i = 0; i < args.length; i++) {
                if (args[i].equals("-showtree")) {
                    showTree = true;
                } else {
                    f = args[i];
                    t = parseFile(f, new BufferedReader(new FileReader(f)));
                }
            }
        } else {
            System.err.println("Usage: java IOScan [-showtree] " +
                              "<directory or file name>");
        }
    } catch (Exception e) {
        System.err.println("exception: "+e);
    }

    return t;
}
```

```java
public static AST parseFile(String f, Reader r) throws Exception {
    try {
        JavaLexer lexer = new JavaLexer(r);
        lexer.setFilename(f);
        lexer.setTokenObjectClass("TokenWithlndex");
        TokenStreamTracker tracker = new TokenStreamTracker(lexer);
        tracker.discard(JavaLexer.WS);  //ignore WS (whitespace)
        JavaRecognizer parser = new JavaRecognizer(tracker);
    }
```
Appendix F (continued)

```java
parser.setFilename(f);
parser compilationUnit();
t = parser.getAST(); // builds AST tree for source file
// call method to build visual representation of tree
doTreeAction(f, parser.getAST(), parser.getTokenNames());
catch (Exception e) {
    System.err.println("parser exception: "+e);
    e.printStackTrace(); // so we can get stack trace
}
System.out.println("AST tree building done");
return t;
}
// build a visual representation of the AST tree in a frame
public static void doTreeAction(String f, AST t, String[] tokenNames) {
    if (t==null) return;
    if (showTree) {
        ((CommonAST)t).setVerboseStringConversion(true, tokenNames);
        ASTFactory factory = new ASTFactory();
        // TreeParser tp = new TreeParser();
        AST r = factory.create(0, "AST ROOT");
        r.setFirstChild(t);
        final ASTFrame frame = new ASTFrame("Java AST", r) ;
        frame.setVisible(true);
        frame.addWindowListener(
            new WindowAdapter() {
                public void windowClosing (WindowEvent e) {
                    frame.setVisible(false); // hide the Frame
                    frame.dispose();
                    System.exit(0);
                }
            } // windowClosing
        ); // addWindowListener
    } // if (showTree)
    JavaTreeParser tparse = new JavaTreeParser();
    try {
        tparse.compilationUnit(t);
    } catch (RecognitionException e) {
        System.err.println(e.getMessage());
        e.printStackTrace();
    } //try
}
// hash java output methods (57 of them in Java 1.3.1)
public static void hashOutputMethods() throws Exception {
    String oNode = ""; // holds each currently read method
    Integer hVal; // used to generate hash code
    BufferedReader inf = new BufferedReader(new FileReader("c:/antlr/outputMethods.txt"));
    while (inf.ready()) {
        oNode = inf.readLine();
        hVal = new Integer(oNode.hashCode());
        outHT.put(hVal, oNode);
    }
    inf.close();
    System.out.println("Hashing of 57 Java output methods complete.");
```
Appendix F (continued)

```java
755     }//hashOutputMethods
756
757     //hash Java loop keywords (for, while, do)
758     public static void hashJavaLoops() throws Exception {
759         String block = "";    //holds each currently read command
760         Integer hVal;     //used to generate hash code
761
762         BufferedReader inf = new BufferedReader(new FileReader("c:/antlr/blocks.txt");
763         while (inf.ready()) {
764             block = inf.readLine();
765             hVal = new Integer(block.hashCode());
766             blockHT.put(hVal, block);
767         }
768         inf.close();
769         System.out.println("Hashing of for, while, do complete.");
770     }//hashJavaLoops
771
772     //hash Java language tokens (163 of them)
773     public static void hashTokens() throws Exception {
774         String token = "";    //holds each currently read token
775         Integer hVal;     //used for hashcode
776         int hV = 0;
777
778         BufferedReader inf = new BufferedReader(new FileReader("c:/antlr/tokens.txt");
779         while (inf.ready()) {
780             token = inf.readLine();
781             hVal = new Integer(Integer.toString(hV));
782             tokensHT.put(token, hVal);
783             hV++;
784         }
785         inf.close();
786         System.out.println("Hashing of 163 Java tokens complete.");
787     }//hashTokens
788
789 }//class IOScan
```
Appendix G: Resources for Information on Semantic Parsing


[5] [http://www.netaxs.com/people/nerp/automata/syllabus.html](http://www.netaxs.com/people/nerp/automata/syllabus.html). Outline and content for a complete course on language grammars, BNF notation, including information on semantic parsing and various kinds of parsers.


[7] [https://javacc.dev.java.net](https://javacc.dev.java.net). The home page for javacc, Sun’s own parser generator for Java.

[8] [https://javacc.dev.java.net/servlets/ProjectDocumentList?folderID=110](https://javacc.dev.java.net/servlets/ProjectDocumentList?folderID=110).

Download site for a variety of grammars that work with javacc (27 languages including C, Java and Visual Basic, Python and Oberon).
[9]  http://home.earthlink.net/~slkpg. Home page for SLK, which claims to be “the only true LL(k) parser” and “the only known near-solution to this NP-complete problem”.

[10]  http://www.gnu.org/software/bison/bison.html. GNU home page for Bison, a parser generator that has been available for UNIX and Linux systems for many years.


[12]  http://cedet.sourceforge.net/info/semantic.html. Semantic: technically a ‘bovinator’ which is a partial lexer. This is a long article that describes ‘bovination’, BNF grammars and how to work with the Semantic product.