

Robin McFarland

Giving Programming Exercises Adaptive Difficulty

Computer Science Tripos – Part II

Homerton College

2017

Proforma

Name: **Robin McFarland**
College: **Homerton College**
Project Title: **Giving Programming Exercises Adaptive Difficulty**
Examination: **Computer Science Tripos – Part II, July 2017**
Word Count: **TBC**
Project Originator: **Mr Michael B. Gale**
Supervisor: **Mr Michael B. Gale**

Original Aims of the Project

Work Completed

Special Difficulties

Declaration

I, Robin McFarland of Homerton College, being a candidate for Part II of the Computer Science Tripos, hereby declare that this dissertation and the work described in it are my own work, unaided except as may be specified below, and that the dissertation does not contain material that has already been used to any substantial extent for a comparable purpose.

Signed [signature]

Date [date]

Contents

1	Introduction	9
1.1	Motivation	9
1.2	Context of the work	9
1.3	Project aims	10
2	Preparation	13
2.1	Starting point	13
2.2	An overview of the project and its components	13
2.3	Detailed discussion	14
2.3.1	Extending a language with blanks	14
2.3.2	Interpreting and pretty-printing the language	14
2.3.3	Encoding representations of programming exercises	15
2.3.4	Setting and adjusting exercises	15
2.3.5	Extension: parsing the language	16
2.3.6	Extension: implementing a GUI	16
2.4	Methodologies in use	17
2.5	Software Requirements Specification	17
3	Implementation	19
3.1	MiniJava's Abstract Syntax Tree	19
3.1.1	Classifications of syntactical constructs	19
3.1.2	Implementing operator precedence and associativity	20
3.1.3	Expression types	22
3.1.4	Degenerate expressions	22
3.1.5	Expressions with multiple possible operators	23
3.1.6	Assignment expressions	23
3.1.7	Local variable declarations	23
3.1.8	Blanks in code	24
3.2	Interpreting the language	24
3.3	Parsing the language	27
3.4	Representing questions and solutions	30
3.5	Adding and removing blanks	33
3.6	The <code>ExerciseSetter</code>	36
3.7	The GUI	38

4	Evaluation	41
4.1	Correct functionality	41
4.1.1	The <code>ExerciseSetter</code> class	41
4.1.2	The parser extension	42
4.1.3	The GUI extension	43
4.2	Trends in difficulty against performance	44
4.3	Development process	47
5	Conclusion	49
5.1	Further work	49
	Bibliography	50
A	The Parser	53
B	Project Proposal	59

List of Figures

3.1	The UML diagram showing the top of the object hierarchy for MiniJASTNodes.	22
3.2	An example of an AST with nodes labeled for clarity.	33
3.3	An example screenshot taken of the GUI peripheral.	38
4.1	A graph showing the contours of the performance function.	42
4.2	Screenshots of the GUI when the student has performed well.	43
4.3	Screenshots of the GUI when the student has performed well.	44
4.4	Increase in measured difficulty with repeated submissions each with a measured performance of one.	45
4.5	This demonstrates the different rates of increase of difficulty found for the different questions. The straight lines are extrapolated past the original points to show that they intersect and thus are not parallel.	46

Acknowledgements

Chapter 1

Introduction

1.1 Motivation

Finding the most effective way to teach young people new skills is an important challenge, impacting educators, parents and learners. Learners want to be delivered new information in an interesting and engaging way that enables them to make use of their new knowledge constructively, and educators want to feel productive as they teach. How well a student learns a new skill is determined by a number of factors. These range from differences between learners (e.g. ability and motivation) to the teaching method used. It is up for debate whether unguided or guided learning styles produce the best results[1], and similarly tasks requiring problem solving are compared with those requiring reciting known solutions[5].

Recently, Computer Science has been included into the national curriculum in the UK for children aged 5 and above. This has raised the question of how best to teach young people a programming language (e.g. Java). The typical method employed in a classroom setting is to give some instruction (which may involve a variety of worked examples) followed by a set of programming exercises for students to complete.

A typical exercise given to students can be conceptualised in the following manner: some unfinished sample code is given to the students, and their task is to fill in the blanks in the code such that the completed code can perform the function required by the question. When students have a firm grasp on the language of choice, it may be reasonable to expect them to write all the code from scratch, that is, no sample code would be given at the start. However, when a student is first learning a language, is less able, or younger, they may find it helpful to have parts of the code provided for them. This enables them to see how the syntax is used and reduces the complexity of the work and understanding required. This method is likely to enhance the student's learning experience and maintain their motivation to continue.

1.2 Context of the work

Attempts have been made before to set exercise difficulty in response to student performance[6]. The author makes use of computerised-adaptive testing (CAT) by de-

veloping a system that enables students to develop their own user profile. The following exercises are then set for students with their difficulty determined by these user profiles. The performance of students is assessed by the students themselves, and thus a measure of their understanding of the subject matter is available directly. The system underwent human trials in 2003 and 2005 with encouraging results, showing increased mean scores in students who used their system versus students who did not. This shows that CAT systems have promise.

1.3 Project aims

A different approach to CAT would be a system that attempts to measure student performance algorithmically, inferring how much the student has understood. The system could then present students with problems varying in complexity as a function of their understanding of the material. The specific questions that the students have to answer would be determined by their teacher and the appropriate curriculum. However, the amount of code that each student is expected to write can be adjusted algorithmically, by adding or removing blanks from a model solution.

This project aims to develop a system that follows this new approach. The system will attempt to teach a programming language by delivering exercises to students for them to solve. The first problem delivered will always be the same, and each student's performance will be measured by the system. Depending on an individual student's performance, the difficulty of the next problem presented to them will be adjusted appropriately. This will be done by changing the amount of the model solution that is initially left blank to fill in. If it is determined that the question is too easy or too hard given the current competence of the student, then the question itself will be changed.

To develop this system, I implement a new language that can incorporate holes in programs and write an interpreter and pretty-printer for it. After this, a representation of programming exercises is encoded, and an algorithm that adds and removes blanks from the model solutions stored in the exercises is written. These tools are brought together in the system itself. This system is able to measure the performance of a student using a heuristic, and use its measurements to adjust the difficulty of programming exercises appropriately. If a student has performed well (as measured by the performance heuristic), then the next exercise presented to them will be more difficult. Conversely, if they have performed poorly, then the next exercise presented to them will be easier.

Extensions to this work include writing a parser for the new language, and implementing a Graphical User Interface (GUI). This GUI demonstrates an example of how a student may interact with the system, by presenting the model solution code in a text box that students can directly edit to fill in the blanks in the solution, and gives feedback on the student's submissions.

The aims of this project are to:

1. Implement a new language, incorporating the concept of blanks.
2. Implement an interpreter and pretty-printer for this language.

3. Encode a representation of programming exercises for the language.
4. Write algorithms that increase and decrease the difficulty of these exercises by adding and removing blanks from the model solution code.
5. Design a system that measures the performance of students as they solve programming exercises, and adjusts the difficulty of the subsequent exercises accordingly.
6. As an extension, parse the language.
7. As another extension, design a GUI to demonstrate how students might interact with the system.

Chapter 2

Preparation

In this chapter, I detail the preparation I undertook before starting on my implementation. I will give my “starting point” in Section 2.1, explaining how much experience I had before work on this project began. In Section 2.2, I will give a brief overview of the project and how its components fit together. Then, in Section 2.4, I will explain the specific software engineering methodologies I employed to ensure implementation went as smoothly as possible. Finally, I will give a detailed requirements analysis in Section 2.3.

2.1 Starting point

I am able to program in Java, the language in which I will be implementing my tools. I have some experience of writing interpreters, primarily of Combinators. I have designed an educational program before as an Extended Project for the AQA Baccalaureate at school. The motivation behind this project was to investigate if simply using a computer could make Key Stage 2 students more motivated to learn maths. The project involved automating the maths tests found in their textbooks by designing templates that were instantiated at run time to provide a greater variety of questions. I have experience designing GUIs, since one was used to present the questions and accept solutions for the Extended Project, and another was an integral part of the Group Project last year.

2.2 An overview of the project and its components

The overall aim of this project is to design a system that can algorithmically adapt the difficulty of programming exercises. Decisions about how to adjust the difficulty are made by referring to measurements of student performance taken as they solve preceding exercises. In order to set programming exercises, we need a language in which a student will solve said exercises. Representations of programs written in this language can be pretty-printed, and are executable by interpreting the code. Programming exercises themselves require representations within the system, and these store representations of the model solution for the exercise, written in the specified language. Wrapped around these exercises is a system for presenting questions, allowing completion of the model solutions, checking validity of submissions, and delivering feedback on submissions. This system

incorporates algorithms that increase and decrease the difficulty of problems by adding and removing blanks from the model solution code. As an extension, a parser is implemented to streamline the functionality of some of these components: the model solution in exercises is defined in a `String` that is then parsed to build the AST representation required, and the system fills blanks by parsing `Strings` containing the code to be substituted with. Finally, as another extension, a GUI is implemented that acts as an interface between the system and both students and teachers, making use of the parser to allow students to directly edit the model solution code as much as they like.

2.3 Detailed discussion

This section will explore the requirements laid out in the Introduction in detail, drawing particular attention to any implementation choices that were made.

2.3.1 Extending a language with blanks

Ideally, I would have preferred not to have to design my own language, but to simply add the concepts of blanks to Java itself. Programs in this extended language would then have to be executable.

An early prototype I built to investigate how I might add blanks to Java involved storing a program as a `String` and executing it as Java code. This seemed a good solution to me as I have experience doing this very thing in Python. Dynamic execution of Python statements and expressions is supported by the `exec` and `eval` functions respectively. These functions even support supplying Python dictionaries containing maps of local and global variables. Unfortunately, doing this in Java is much more complex, requiring the use of external tools. Bearing in mind that these external tools have no inbuilt support for blanks in code, I would have to write some potentially very substantial extensions to these tools, and these factors combined persuaded me to follow an alternate path.

I decided to design my own language that would very closely imitate Java, but include the concept of blanks in code within the grammar itself. This language would imitate Java in order to demonstrate that Java itself, a language taught to Part IA students in the Computer Lab and many other beginners, could be taught using my system. However, re-implementing all of Java would have been too much unnecessary complexity, since this was not the focus of the project. I thus designed my language, which I call MiniJava, to only imitate a small subset of the Java grammar [2, p.714].

2.3.2 Interpreting and pretty-printing the language

Representations of programs written in MiniJava should be executable, and should also be able to be pretty-printed to allow a user to read the code intuitively as text. As an extension, it would be nice to be able to take a textual version of a program and convert it into an AST representation in my implementation by parsing it.

To interpret MiniJava, variables used throughout the program must be managed properly. This involves allowing the declaration of new variables, managing their scope, and

updating their values as they are assigned to. The most effective methods for managing variables as the interpreter runs are to store them in symbol tables or in stack frames. Early prototypes showed that storing variables in symbol tables made the management of scope more complex, as the “depth” at which a variable is declared must be explicitly stored too, so that the variable can go out of scope at the appropriate time. When variables are stored in stack frames, scope management becomes implicit within the structure of the stack, so I decided to use this method.

2.3.3 Encoding representations of programming exercises

A programming exercise consists of a question, represented by a `String`, and a model solution, represented by a MiniJava AST. An exercise allows the blanks in its solution to be filled in by students, and complete solutions are executable. Exercises record constraints that must be satisfied by any complete solution to be declared valid. They also provide methods to make them easier or harder by introducing blanks to or removing them from the model solution prior to the question being presented. The algorithm I designed to do this is described in Section 3.5, and takes inspiration from both the standard Depth First Search algorithm and the more obscure Reverse Level Order Traversal. Designing this algorithm was a significant challenge during my preparation, as there are a lot of properties that it had to have.

A representation of an exercise provides some degree of variety, such that an exercise would not always ask the question, “What is the square of 5?”, for example, but rather “What is the square of n ?”, where n is a random positive integer between some minimum and maximum values¹.

2.3.4 Setting and adjusting exercises

An exercise is set by first delivering its respective question, and then presenting the incomplete model solution to the student. The blanks in this solution should be obvious, and they also should be editable. In the case that students can not directly modify the code in its entirety, this will require labelling each blank with a unique identifier, so that students can select the correct blank to fill. Alternatively, if students can directly modify the model solution, then these identifiers will not need to be shown. Once the solution has been completed and submitted, the solution must be run so that whether or not it solves the problem can be decided. This will require interpreting the code, since there are infinitely many ways of implementing a correct solution. Either way, some feedback needs to be delivered to the student, either telling them that their solution was correct, or that they have made a mistake.

Once a solution has been validated and shown to be correct, the performance of the student must be measured. The information most readily available to us with which to derive this measurement is the number of attempts a student required to produce this correct solution, and the number of nodes in their solution’s AST compared with the

¹Note that I am not asking students to solve the problem “write a routine that squares any number”, but simply randomising the specific number that they are asked to square.

model solution’s AST. The more attempts a student required, the worse their performance, and I would argue that using many more nodes than is necessary also signifies a lack of understanding, and thus results in a lower performance.

Once we know how well the student performed we can decide how difficult their next exercise should be to best match their competence. If it is determined that a student has “mastered” the current question, then they will be presented with a more difficult question to answer. If this is not the case, then the amount of the model solution that is blank will be changed instead.

2.3.5 Extension: parsing the language

I kept the goal of writing a parser of the language as an extension to this project, as a parser is not necessary to demonstrate the power of the other tools being written. I decided that if I did implement a parser I would use a parser generator, as writing a parser myself from scratch would be time consuming and error prone. Of the available parser generators, I investigated ANTLR[3] and JavaCC², as they both use Java and are both popular. Both of these parser generators support tree building (which is necessary for building the AST representation of a program in my implementation of the language), however JavaCC provides this as a separate tool while ANTLR provides the functionality by default. Looking at how the two tools are used to generate parsers for very simple grammars, I found ANTLR’s work flow much more amenable. ANTLR also has more high profile users, like IntelliJ IDEA, Groovy, Twitter and Numbers (Apple’s spreadsheet program). Based on this high recommendation, and my own opinion that it is more intuitive to use, I decided to use ANTLR if I came to implement this extension to my project.

2.3.6 Extension: implementing a GUI

The GUI should act as an interface between the system and both teachers and students. Teachers should be able to manually adjust the difficulty of a given question. Students meanwhile should be shown the question itself alongside the code representation of the model solution requiring completion. On the condition that a parser has been implemented, the GUI should allow direct editing of the model solution, the result of which can then be parsed and set as the solution to the exercise to allow validation. The most popular frameworks for implementing GUIs in Java are Swing³ and JavaFX⁴, since they have the most support. Although I have more experience working with JavaFX from the group project, I was not the person who did the work on the initial set up. Also, the Swing plugin for IntelliJ (the IDE that I am using) has a drag and drop interface for designing UIs, and greatly improves the experience of working with Swing, so I decided to use this if I was able to work on the extension.

²<https://javacc.org/> - last accessed 3rd May, 2017

³<https://docs.oracle.com/javase/7/docs/api/javafx/swing/package-summary.html>

⁴<http://www.oracle.com/technetwork/java/javafx/documentation/reference-137773.html>

2.4 Methodologies in use

In order to ensure correct functionality of my system and its components, the implementation process proceeded under the influence of Test Driven Development (TDD). This required writing unit tests for code that had not yet been written. These tests were designed with respect to the requirements, such that any piece of software passing the tests also satisfies the requirements. It was shown that these tests failed before the unit was implemented, and that they are passed after. These tests were then kept to act as regression tests.

2.5 Software Requirements Specification

The following requirements are annotated with the appropriate identifiers from the MoSCoW method of prioritising requirements⁵.

1. The language in which a student should solve exercises will be made to support blanks. **M**
2. This language will be interpreted and pretty-printed. **M**
3. This language will also be parsed. **C**
4. Representations of programming exercises will be implemented. **M**
5. These exercises will have components that are randomly generated to provide variety. **C**
6. A system that sets exercises will be developed to measure the performance of a student and adjust the difficulty of following exercises. **M**
7. A GUI will be implemented to demonstrate how users might interface with the system. **C**

This chapter has described the preparation undertaken before any code was written, including analysing requirements, deciding which libraries to use, and choosing software design methodologies to follow.

⁵<http://research.omicsgroup.org/index.php/MoSCoW> - last accessed 10th May

Chapter 3

Implementation

In this chapter I detail the work completed. I begin by describing how I implemented MiniJava in Section 3.1, as well as an interpreter and a parser for MiniJava’s grammar in Sections 3.2 and 3.3. I then describe how I encoded a representation of a programming exercise solved using MiniJava, and give an example of the encoding of a specific exercise in Section 3.4. In Section 3.5, I describe the algorithm that allows the addition and removal of “blanks” within a MiniJava program. In Section 3.6, I describe how I developed the system that sets and adjusts the exercises, the `ExerciseSetter` class, and in Section 3.7 I describe the implementation of the Graphical User Interface (GUI) as an extension.

3.1 MiniJava’s Abstract Syntax Tree

In this section I describe the implementation of the MiniJava language. This language was designed to be similar to a small subset of Java, and is the language in which students complete the exercises delivered by my system. I have incorporated the concept of blanks in programs into its grammar, to make adding blanks to the model solutions of exercises much easier.

3.1.1 Classifications of syntactical constructs

Each class in my implementation represents a different production rule in the abstract syntax for MiniJava, found in Appendix A in the form of a parser specification. In order to allow other parts of my code to operate on instances of these classes without needing to know exactly what they are, every class in the implementation implements the `MiniJASTNode` interface. This acts as the top of the object hierarchy for my implementation.

It is important to differentiate between MiniJava expressions and MiniJava statement as they must be used in different places. For example, in the production rule `WHILE parExpression statement` representing a `while` loop, `parExpression` may be substituted with any MiniJava expression (surrounded by parentheses), while `statement` may be substituted with any MiniJava statement. The reverse is not allowed however, meaning that `statement` may not be replaced by a MiniJava expression and vice versa. This distinction is made by the interfaces `Expression` and `Statement` (which all

MiniJava expressions and statements implement respectively), and the abstract classes `ExpressionBase` and `StatementBase` (which all MiniJava expressions and statements extend respectively). Both interfaces and abstract class are used here because of the need to include blanks in the grammar. There are two classes representing blanks: `FillableBlankExpr` and `FillableBlankStmnt`. The first of these should be a MiniJava expression and the second should be a MiniJava statement, meaning they should extend `ExpressionBase` and `StatementBase` respectively, but both of them must also extend the abstract class `FillableBlank` (this is explained on page 24). Since classes in Java cannot extend more than one base class, the only alternative is for them to implement the `Expression` and `Statement` interfaces. However, removing `ExpressionBase` and `StatementBase` would result in too much repeated code across the whole grammar, so the grammar requires both the interfaces and the abstract classes. The enumeration `SingleWordStmnt` (each value of which represents a MiniJava statement that is trivial, consisting of zero or one word) must also be a MiniJava statement, and since enumerations cannot extend classes, it must implement `Statement`, further cementing the need for both the interfaces and the abstract classes. Since the distinction between expressions and statements is the most high level distinction one can make, both the `Expression` and `Statement` interfaces extend the `MiniJASTNode` interface, ensuring that every class in the implementation also implements it.

As important as the distinction between expressions and statements is, there do exist some expressions that can be used in statements. For an example, the MiniJava code `i = 5;` is a statement because it ends in a semicolon, “;”, but this statement consists entirely of the expression `i = 5` (this allows for the chaining of assignments found in the Java grammar [2, p.589]). This sort of expression within a statement construct is not possible with all expressions, for example the code `i + 4;` would not be valid MiniJava. To make the distinction between expressions which may be used in statements like this and those which may not, I made the interface `StatementExpression`. This interface is only implemented by those expressions which can be used in a statement this way.

3.1.2 Implementing operator precedence and associativity

A feature of MiniJava is the concept of operator precedence, which is used to disambiguate an otherwise ambiguous grammar of expressions. Without the concept of operator precedence, the code `1 + 2 * 3` is ambiguous, as it can be read as either $(1 + 2) * 3$ which equates to 9, or as $1 + (2 * 3)$ which equates to 7. Since a single arithmetic expression cannot have more than one value, a decision must be made as to which of these interpretations is declared valid. The decision taken aligns with the standard arithmetic order of operations, such that the answer of 7 is the correct answer in this case. This example demonstrates that the multiplication operator, “*”, has a higher precedence than the addition operator, “+”. All the possible operators in MiniJava are found within this precedence hierarchy, and I needed some way to encode this. The method I chose for this is the natural choice, since it is suggested by the Java grammar itself [2, p.723]. The production rule for `AdditiveExpression` is as follows:

AdditiveExpression:

MultiplicativeExpression (1)

AdditiveExpression + MultiplicativeExpression (2)

AdditiveExpression - MultiplicativeExpression (3)

and the production rule for MultiplicativeExpression is as follows:

MultiplicativeExpression:

UnaryExpression

MultiplicativeExpression * UnaryExpression

MultiplicativeExpression / UnaryExpression

// Note that the modulus operator, %, was removed from MiniJava

This means that for any expression $A + B$, the expression A must only contain subexpressions that use operators with precedence at least as high as $+$ and $-$, and B must only contain subexpressions that use operators with precedence higher than $+$ and $-$. For any expression $A * B$, neither A nor B can contain subexpressions that use operators with precedence lower than $*$ and $/$. This in turn indicates that an expression such as $1 + 2 * 3 + 4$ can only be interpreted as $1 + (2 * 3) + 4$, and an expression such as $1 * 2 + 3 * 4$ can only be interpreted as $(1 * 2) + (3 * 4)$. I wanted an operator precedence for MiniJava that is hardcoded within the grammar of MiniJava itself, just as the operator precedence of Java is hardcoded within the grammar of Java. The production rules for Java shown above suggest a natural implementation of the class that represents addition expressions:

```
class AddExpr {
    private boolean isPlus;
    private AddExpr leftSide;
    private MultExpr rightSide;
    ...
}
```

Here, the `isPlus` field records whether the expression is an addition or a subtraction, `leftSide` and `rightSide` store the left and right operands of the expression, and `MultExpr` is the class representing multiplication expressions. What has been implemented so far encapsulates parts (2) and (3) of the `AdditiveExpression` production rule above, but not part (1). This part is encapsulated by making the class `MultExpr` extend the class `AddExpr`, such that an instance of `MultExpr` may always be used in place of an instance of `AddExpr`. Note that since multiplication operators are higher in the operator precedence hierarchy, their representations are lower in the object hierarchy. This then is the solution to the operator precedence problem: to enforce an operator precedence hierarchy, first enforce an inverse object hierarchy in their representative classes, and then choose the type of each expression's operands to match the relevant terminal symbols in the relevant production rule.¹

¹Note that if you were to look at my implementation of `AddExpr`, `leftSide` and `rightSide` would both be `ints`, not `AddExpr` and `MultExpr` instances. The reason behind this is explained in section 3.5.

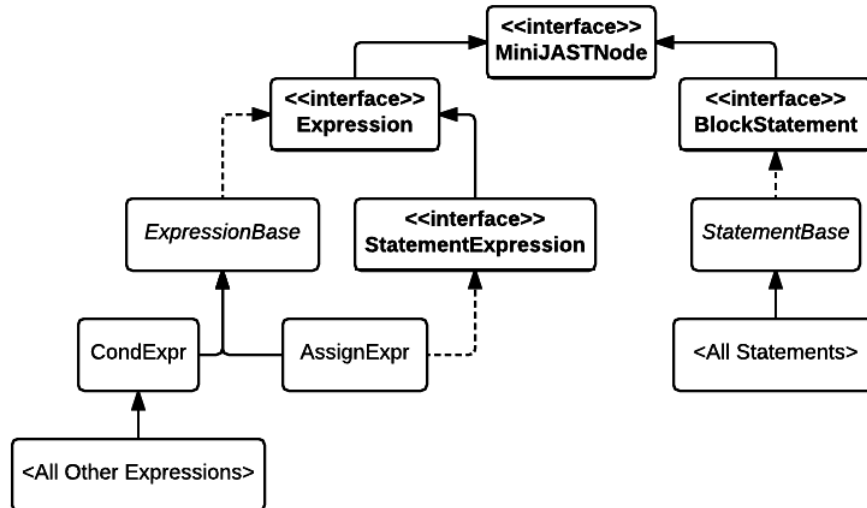


Figure 3.1: The UML diagram showing the top of the object hierarchy for MiniJASTNodes.

Figure 3.1 shows the top of the object hierarchy for the implementation of MiniJava. It demonstrates the distinction between MiniJava expressions and statements by way of the **Expression** and **Statement** interfaces, the extra classification of expressions that can be used as statements given by **StatementExpression**, and shows the beginnings of the operator precedence hierarchy hardcoded into the expressions.

Note also that as a side-effect of this implementation, the ambiguous code `1 + 2 + 3` also only has one possible interpretation, namely `(1 + 2) + 3`. This follows the established rules of operator associativity, which are now also hardcoded into the grammar.

3.1.3 Expression types

In MiniJava, every expression has a type which consists of a primitive type and may or may not be an array type. There are four possible primitive types: `boolean`, `char`, `int`, and `double`. The primitive types are represented in the grammar by the enumeration, `PrimType`. The type of an expression then is represented by the class `Type`, an instance of which stores the appropriate `PrimType` value and a flag determining whether or not it is an array type. These `Types` are used by the interpreter to determine whether or not variables can contain certain values, and whether operators are being supplied with the appropriate operands.

3.1.4 Degenerate expressions

The majority of statements in MiniJava are composed of statements or expressions, but there are some that consist of only a single word, or even no words at all. To have an entire class representing each of these trivial statements seems wasteful, so the enumeration `SingleWordStmnt` represents four such statements: the break statement (`break;`), the continue statement (`continue;`), the return statement (`return;`), and the empty statement (`;`). Although there are no functions in the MiniJava grammar, the return statement can be used to immediately halt evaluation.

3.1.5 Expressions with multiple possible operators

As we have seen previously, `AddExpr` defines the boolean field `isPlus`, to determine whether this instance represents an addition or a subtraction. There are some expressions where the number of possible operators that can be used in that expression is greater than two, in which case a boolean field will not be enough to disambiguate. `RelationExpr` and `AssignExpr` are the two expressions for which this is the case, since there are four relation operators and five assignment operators. These possible operators are represented by the two enumerations `RelationOp` and `AssignOp`. Thus `RelationExpr` stores a value of `RelationOp`, and `AssignExpr` stores a value of `AssignOp`.

3.1.6 Assignment expressions

There are two possible MiniJava expressions that can be assigned to, i.e. they can appear on the left hand side of an assignment expression. These two, identifiers and array accesses, are represented by the classes `Id` and `ArrayAccess`. To signify that they can both be assigned to, both classes implement the `AssignLHS` interface, an interface used for only this purpose. Similarly, there are two possible MiniJava expressions that can be assigned to an array type variable. These are array creation expressions (for example `new int[4]`) and array initialisation expressions (for example `{1, 2, 3, 4}`), which are represented by the classes `arrayCreation` and `ArrayInit` respectively. To signify that both can be assigned to arrays, they both extend the abstract class, `ArrayAssignRightSide`. To justify why `AssignLHS` is an interface and `ArrayAssignRightSide` is an abstract class, we consider the difference between the expressions that can appear as subexpressions of other expressions, and those that cannot. Identifiers and array accesses may appear as a subexpression of some other expression. Because of this, the index operator used in array accesses must appear within the operator precedence hierarchy described previously, and so too must identifiers. If identifiers were not present in the operator precedence hierarchy, then they could never be substituted for the terminal symbols in the production rules seen previously, ruling out valid expressions like `i + j`. If identifiers and array accesses must appear in the operator precedence hierarchy, and high up in it too, then that means their representative classes must extend some class representing an operator lower down in the hierarchy. Thus `AssignLHS` must be an interface and not an abstract class, as `Id` and `ArrayAccess` already subclass some other class (namely `UnaryExpr`). Array creation and initialisation expressions however can never appear as a subexpression of some expression (since only single-dimensional arrays are allowed, expressions such as `{{1},{2,3,4}}` are not valid). This means that they do not appear in the operator precedence hierarchy, and thus their representations do not need to extend any base class. Thus `ArrayAssignRightSide` can usefully be an abstract class, reducing the amount of repeated code in these two classes.

3.1.7 Local variable declarations

MiniJava uses local variable declarations (for example, `int i = 4, ar[];`) in the same way that Java does: to declare and initialise new variables. Local variable declarations can be thought of as being made up of a primitive type and a list of so called “variable

declarators”, which are identifiers that may or may not be followed by square brackets to show the new variable is an array, and which may or may not be followed by an equals sign, “=”, and a value with which to initialise the new variable. My implementation represents local variable declarations using the class `LocalVarDec`, which consists of a `PrimType` and a list of `VarDeclarator` instances. The `VarDeclarator` class stores a `String` containing the name of the variable, flags to determine whether the variable is an array type and whether an initialiser expression is given, and the initialiser expression itself if there is one.

3.1.8 Blanks in code

What distinguishes MiniJava from other languages is that the concept of a blank, a missing piece of code, is built into the grammar itself. Representations of blanks in code should enable a user to fill one with their own code. To do this programmatically, it is necessary for representations of blanks to each come with a unique id, such that users can choose which blank they are filling and the blank be searchable within the AST representing the program. The abstract class `FillableBlank` fulfils this requirement by storing a `static` reference to an `AtomicInteger` and a private `int` field, `id`. Whenever a new blank needs to be represented in an AST, an instance of a class that extends `FillableBlank` can be supplied with a unique identifier that has not yet been used in this runtime environment, by using the `AtomicInteger`’s `incrementAndGet` method and storing the result in `id`. The two concrete classes that extend `FillableBlank` are `FillableBlankExpr` and `FillableBlankStmnt`. The first of these can be used to replace a MiniJava expression (and thus implements the `Expression` interface so that it can be stored in a MiniJava AST), and the second can replace a MiniJava statement (and thus implements `Statement`). `FillableBlankExpr` stores an `Expression` field, `studentExpr`, which, if not `null` is representative of the code with which a user has filled the blank. Similarly, `FillableBlankStmnt` stores an `Statement` field, `studentStmnt`.

In this section I have described the implementation of my MiniJava language.

3.2 Interpreting the language

In this section I explain how MiniJava programs are interpreted.

To correctly interpret MiniJava, several requirements must be met by the interpreter:

1. The scope of variables must be handled, such that when a variable is used its current value can be accessed, and the same variable can’t be declared twice in one scope.
2. Expressions must be able to propagate their values upwards, so that surrounding expressions and statements can use them.
3. Flow control must be handled, such that `break` and `continue` statements can be used in loops.
4. If an exception is raised during execution, the exception should propagate up and halt execution, preferably delivering a useful message.

The scope of variables is handled by the interpreter with instances of the `Context` class. Part of its definition is shown below:

```
public class Context {  
    public Stack<HashMap<String, Type>> namesToTypes;  
    public Stack<HashMap<String, Object>> namesToValues;  
    ...  
}
```

The `namesToTypes` field is a stack of maps from names (as `String` values) to instances of my `Type` class. Every variable that is declared in this scope will have an entry in the map on top of this stack, recording the type of that variable. The `namesToValues` field is a stack of maps from names to instances of `Object`. If a variable has an entry in the map on top of this stack, then the variable has the value stored in the map with it in this scope. It is possible for a variable to have an entry in the map on top of the `namesToTypes` stack, but not in the map on top of the `namesToValues` stack: this indicates that the variable has been declared but not initialised in this scope.

During execution, the interpreter must keep track when variables will go out of scope. This is handled by the `stepIn` and `stepOut` methods in `StatementBase`. The `stepIn` method is called whenever execution enters a new, deeper, scope. This method copies every key value pair in the maps on top of their respective stacks into new maps, and pushes these new maps onto the stack. Thus every entry in one of the stacks represents a different scope: as you move down the stack, you move outward through the nested scope. The `stepOut` method is called whenever execution leaves the current scope. At this point, all the variables that do not exist in the new scope should be forgotten, and every variable that does exist in this new scope should have their value updated to the value it was in the old scope. The `stepOut` method first pops the top off both stacks, storing the top of the `namesToValues` stack in the local variable `oldMap`. Then, for every key in `oldMap`, if this key also appears in the new top of the `namesToTypes` stack, the corresponding value in `oldMap` is used to update the top of the `namesToValues` stack. In this way, variables can be updated in deeper levels of scope, and retain their new values when execution moves up the nested scope.

Consider the expression $(2*3)+4$. In MiniJava, this expression might be represented by a `AddExpr` storing a literal with the value 4 as the right operand, and an `MultiExpr` as the left operand. This `MultiExpr` would store literals with values 2 and 3 as the left and right operands. During the evaluation of an expression, the most deeply nested parts of the expression must be evaluated first, so that their values can be used further up in the nested expression. There is no point trying to evaluate the `AddExpr` before we know the value of $(2*3)$. Thus, the `MultiExpr` needs a way to communicate to the `AddExpr` that it has a value of 6, so that the `AddExpr` can be evaluated to 10. Likewise, whatever context the `AddExpr` is in needs to be told it has the value 10. The way this is done in MiniJava is using implementations of the `ReturnValues` abstract class. This class represents the notion that some value is being returned by an expression, but allows any type of value to be returned, no matter what primitive type the value takes, and whether it's an array or not. To return a value of a particular primitive type, the appropriate one

of these four implementations must be used: `ReturnValuesBool`, `ReturnValuesChar`, `ReturnValuesInt`, or `ReturnValuesDouble`. Each of these has a public `value` field of the appropriate type. If an array value is being returned by an expression, then a further implementation of `ReturnValues` must be used, the generic class `ReturnValuesArray<T>`. This class makes use of Java Generics as it stores an internal `ArrayList<T>` of values.

In the same way that MiniJava expressions dispense values, MiniJava statements dispense control flow commands. These commands can be to “break”, to “continue”, to “return”, or there can be no command, in which case execution continues as normal. The MiniJava interpreter represents these commands using the `FlowControl` enumeration, which has the four values `BREAK`, `CONTINUE`, `RETURN`, and `NONE`. Whenever a statement is executed, the interpreter looks for which value is dispensed and reacts accordingly.

Several custom exceptions are used by the interpreter, to differentiate between when an error occurs in the MiniJava program being interpreted, and when an error occurs in the interpreter itself. All these custom exceptions subclass the base class of `MiniJavaException`, and the majority of them are analogous to standard Java exceptions. For example, the `OutOfBoundsException` is thrown by the interpreter whenever an index is used in an array where no such index exists. This exception is analogous to Java’s `IndexOutOfBoundsException`.

The MiniJava interpreter is implemented using the `evaluate` method in the `Expression` interface, and the `execute` method in the `Statement` interface. The method signatures of these two methods are shown below:

```
public interface Expression extends MiniJASTNode {
    ReturnValues evaluate(Context c) throws MiniJASTException;
    ...
}

public interface Statement extends MiniJASTNode {
    FlowControl execute(Context c) throws MiniJASTException;
    ...
}
```

It can be seen that the interpreter makes use of all the components meeting the requirements for the interpreter. The `evaluate` method is overridden by every class representing a MiniJava expression, and the `execute` method is overridden by every class representing a MiniJava statement. This means that both methods are recursive: expressions being evaluated will call the `evaluate` method on their subexpressions for example. Typically, execution is initiated with an empty `Context` object that becomes more populated as more variables are declared. This interpreter can also be plugged in to peripherals. The values of variables in the outermost scope can be accessed after execution has terminated, and MiniJava has an inbuilt print statement, represented by the class `PrintStatement`, that can print the values of expressions to a file. Peripherals could also, if they wanted, make use of the custom exceptions thrown during execution, and the initial `Context` object.

It is important to remember that the MiniJava grammar also contains the notion of blanks in code, and thus the interpreter must be equipped to deal with these appropriately.

If a blank has not yet been filled, then the interpreter should throw a custom exception, but if the blank has been filled, then execution must continue as if there was no blank there. The `FillableBlankExpr` and `FillableBlankStmt` classes override the `evaluate` and `execute` methods in similar ways to fulfil these requirements. If the `Expression` or `Statement` field representing student code is discovered to be `null` at evaluation or execution, then the custom exception `BlankEmptyException` is thrown. If, however, the field is not `null`, then the blank simply acts as a conduit to pass information to and from the student code. `FillableBlankExpr` evaluates the student expression with the `Context` object supplies, and simply passes the returned `ReturnValues` object upwards as if it were itself an evaluated expression, while `FillableBlankStmt` does the same for statements except with the returned `FlowControl` object.

In this section I explained how MiniJava is interpreted.

3.3 Parsing the language

In this section I explain how a parser for MiniJava was implemented using the ANTLR parser generator [3], and how a representation of an input MiniJava program can be built using it.

The parser grammar (which can be found in Appendix A) is based on an existing ANTLR parser grammar for Java². It would not have been worthwhile for me to write an entire grammar from scratch when there was such a readily available alternative. It is always good to use existing libraries or tools if they are available.

The ANTLR tool generates recursive descent parsers from input grammars. The typical usage of ANTLR follows this process:

1. Write a grammar in a grammar file.
2. Run the ANTLR tool on this grammar file to generate a parser.
3. When this parser is executed with input source code, ANTLR's internal representation of the corresponding parse tree is built.
4. Using either a "listener" or a "visitor", walk the generated parse tree, taking actions at each node, processing the source code as desired.

The following code shows how to build a MiniJava representation for the source code `i = 4;`.

```

1 AssignExpr aE = new AssignExpr();
2 Id i = new Id();
3 i.setUpId("i");
4 Literal four = new Literal();
5 four.setUpLiteral(PrimType.INT, "4");
6 aE.setUpAssignExpr(i, AssignOp.EQ, four);
7 ExpressionStmt eS = new ExpressionStmt(aE);

```

²<https://github.com/antlr/grammars-v4>

Considering this representation as a parse tree, this can be seen as line 0 (i.e. the invisible line before line 1) visiting the `ExpressionStmnt` node and visiting its only child, the `AssignExpr` node in line 1. This node's first child is visited in lines 2 and 3, and its second child is visited in lines 4 and 5. Once all the children of the `AssignExpr` have been visited, the instance itself can be set up on line 6. Once all the `ExpressionStmnt`'s children have been visited, its instance can be set up on line 7. This example demonstrates that in general, program representations are generated by recursively visiting each expression or statement's children, in a similar way to how a program representation is interpreted.

By default, the ANTLR tool generates a listener interface and base class along with the parser. A listener is characterised by providing callback methods that are triggered by an automatic parse tree walker: listener methods don't have to explicitly visit their children. A visitor on the other hand gives the user more control over the walk, requiring explicit commands to visit the children of each node. While this means that it takes less work to use a listener (as the parse tree walker is already built in), the visitor paradigm is more appropriate for building my library's representation of the MiniJava source code. While the listener methods all return `void`, requiring the instances representing the children nodes to be stored somewhere in memory prior to setting up the current node, the base visitor, `MiniJavaBaseVisitor<T>`, is generic, with method signatures such as `public T visitBlock`. This means that the visitor methods can be made to return `MiniJASTNodes`, which can be cast to `Expressions` or `Statements` as appropriate (as shown in Figure 3.1 on page 22). By using a visitor rather than a listener, I was able to more closely imitate the way (shown above) that program representations are build in code. My visitor implementation can be found in `MiniJavaASTBuilder`, a class which extends the `MiniJavaBaseVisitor<MiniJASTNode>` class supplied by the ANTLR tool.

Java, and by extension MiniJava, is a language that suffers from the “dangling-else” problem: a particular problem arising from an ambiguous grammar. Consider the MiniJava code shown below:

```

if (i == true)           (1)
    if (j == false)      (2)
        k = 1;
else
    k = 2;

```

A parser could not make a decision as to whether the `else` statement “belongs” to the first or the second `if` statement. Since the grammar is ambiguous, a decision must be made for the parser. The decision taken in Java[2, p.417], and thus in MiniJava, is to attach the `else` statement to most inner `if` statement possible (which in the example above would be the one labeled (2), even though the indentation may suggest that the programmer intended it to be the one labeled (1)). To implement this decision, the grammar distinguishes between `Statements` and `StatementNSIs` (where NSI stands for “No Short If”, i.e. no `if` without `else`). The way this solves the problem becomes apparent on inspection of the following rules in the grammar (presented in context in Appendix A):

```

statement:
    ...
    | IF parExpression statement
    | IF parExpression statementNSI ELSE statement
    ...
    ;

```

```

statementNSI:
    ...
    | IF parExpression statementNSI ELSE statementNSI
    ...
    ;

```

The `statement` rule gives productions for both a short `if` (i.e. with no attached `else`) and a long `if`, while the `statementNSI` rule only produces a long `if`. It can be seen that using these rules to parse the code above will reach a dead end if the first `if` statement is parsed as a long `if`, since its first sub-statement would have to be a `statementNSI`, and this rule has no viable way of parsing the second `if` statement. Of course, the visitor rules for `statements` and `statementNSIs` make no distinction as they build the MiniJava representation, returning just a standard short or long `if` representation as appropriate.

Another problem encountered while writing the parser can be discovered on inspection of the local variable declaration production shown below:

```

blockStatement:
    primitiveType variableDeclarators SEMI    # localVariableDeclaration
    ...
    ;

```

This production is used in the creation of `LocalVarDec` instances, which require the appropriate `PrimType` to set up correctly. As the visitor arrives at a node produced like this in the parse tree (by calling the `visitLocalVariableDeclaration` method found in `MiniJavaASTBuilder`) it needs to visit the node's children in order to gather the necessary objects to set up the `LocalVarDec` object. Since visitor methods return `MiniJASTNodes`, the `visitPrimitiveType` method (also found in `MiniJavaASTBuilder`) needs to return an object whose class implements `MiniJASTNode` and represents a primitive type. Until this point, no such class existed. The enumeration `PrimType` does not implement `MiniJASTNode`, as that would require overriding all the methods declared in the interface separately for each value, as they are in `SingleWordStmnt`. Since this would be a lot of wasted code, I instead wrote a new class, `PrimTypeCarrier`, for precisely this purpose. This class implements `MiniJASTNode` and has a public field of type `PrimType`, meaning it only needs to implement (trivially) all of the interface methods once. An instance of this class can thus be created in the `visitPrimitiveType` method (setting the `PrimType` field value appropriately), which can then be returned to the `visitLocalVariableDeclaration` method for setting up the `LocalVarDec` object.

In this section I have described how the ANTLR parser generator can be used to generate parsers for the MiniJava grammar, and how program representations can be built using visitors.

3.4 Representing questions and solutions

In this section I explain how the general notion of a question is represented using the `AbstractPExercise` class, and give an example of how specific questions can be encoded by extending this class.

When presented to a student learning a programming language, a “question” can be broken up into the following components:

- A description of what the student is expected to do.
- A model solution to the problem presented with parts of the code left blank for students to fill in.
- A measure of difficulty so that both teaching staff and students can monitor progress and assess performance.

Following these guidelines, a representation of a programming question should be capable of the following:

1. Storing the description of what the student is expected to do.
2. Building a MiniJava representation of the model solution.
3. Adding and removing blanks from the solution to make the problem harder or easier.
4. Calculating the difficulty of the current problem, given how difficult the question is to answer and how much of the model solution is blank.
5. Filling those blanks with MiniJava representations of student submitted code.
6. Executing the solution.
7. Checking that the problem is solved by the provided solution.

The abstract class `AbstractPExercise` provides the framework for classes extending it to fulfil all these requirements. I explain how the points above are addressed by `AbstractPExercise` in order.

1. Storing the description of what the student is expected to do

The `String` field `question` stores the question provided to students. This field is initialised in the constructor. In order to present the model solution along with the question, MiniJava program representations need to be converted back in to `String` representations. To implement this, the `stringRepr` method was added to both the `Expression`

and `Statement` interfaces, with the latter's version taking an integer argument representing how many blocks deep the statement is. This method is implemented in every MiniJava expression and statement by recursively combining String representations of the sub-expressions and sub-statements, effectively turning the representation in the abstract syntax into one in the concrete syntax. Depending on how blanks are filled by the specific peripheral in use, it may be useful to print blanks with their unique ids or without. The `FillableBlank` base class records with a `static final boolean` flag whether ids should be shown or not. If they should, then the blank with id n is printed: `..(n)..`, otherwise it is printed `.....`.

2. Building a MiniJava representation of the solution

Since the solution to the exercise needs to be executed (to ensure that it is correct), it must be stored as a representation of the MiniJava code. This representation is built by the parser using the `String` field `solutionCode` as input. The representation is stored in the `Statement` field `solution` of the `AbstractPEExercise` class.

3. Adding and removing blanks from the solution

The algorithms used to add and remove blanks from the solution code are complex enough to warrant their own section, and are thus discussed in section 3.5.

4. Calculating the difficulty of the current problem

As already stated, the difficulty of a question is related to how challenging it is to answer the question, and how much of the model solution is blank. Research into pedagogy was beyond the scope of this project, and it is expected that teaching staff would have a better understanding of how difficult different exercises would be for their own students anyway, so for now the difficulty measurement system is very much a place-holder. An integer is supplied to the base constructor by every implementation indicating the supposed difficulty of that question, with higher numbers assumed to indicate harder questions. This integer is combined with a float measuring the percentage of the model solution that is currently blank to give a measure of the difficulty of the overall problem. This means that for a question with difficulty n , the difficulty of the associated problem can range from n to $n + 1$.

To work out how much of a MiniJava AST is blank, we need to count how many nodes in this tree have been replaced by a blank, and divide that by the total number of nodes in the tree. In order to calculate how many nodes in a tree have been replaced with a blank, we need to know how many nodes have been replaced by each *specific* blank. As such, the `FillableBlank` base class stores an integer, supplied at a blank's construction, that records how many nodes this blank replaced. The standard way to count the number of nodes in a tree which has root N is to count the number of nodes in each subtree that has a root that is one of N 's children, and add one to the total. Since we are counting not only the total number of nodes in the tree, but also how many of those nodes are blank, the method used to count the nodes must return both these

numbers simultaneously. This is done using the class `NodeCount`, which stores two integer fields, one, `empty`, records the number of blank nodes in the tree, and the other, `filled`, records the number of nodes which are not blank. The `getTreeSize` method returns a `NodeCount` object, and is implemented by every class in the MiniJava implementation. It uses the recursive approach to counting nodes in a tree, distinguishing between nodes which are blank and nodes which aren't. To work out how much of the AST is blank, this method is called on the root of the tree so that the returned `NodeCount` object can be used to calculate the percentage using the formula `empty / (empty + filled)`.

5. Filling blanks with student submitted code

Since representations of blanks in code have their own unique identifier, the blank that a user wishes to fill can be selected using it. To fill a blank with the representation of a user's code, the representation of the blank itself must be found within the program's AST, before the representation of the code is stored within it. This functionality is found within the `fillBlank` method of the `AbstractPExercise` class. This method takes the identifier of the blank to be filled and the AST of the code to fill it with. The method then uses a depth-first-search through the AST to find blanks which, when found, have their identifiers compared with the target identifier. If they do not match, then the search moves on, but if they do match, then it is determined whether the blank is an instance of `FillableBlankExpr` or `FillableBlankStmnt`. If it is the former, then the blanks `studentExpr` field is set to the user's AST. If it is the latter, then it is the `studentStmnt` field that is set. If the search fails, then it can be reported that no blank with the suggested identifier exists within the solution.

6. Executing the solution

Since the `AbstractPExercise` class stores the AST of the solution, executing the solution is as simple as using the inbuilt interpreter on it. The `runSolution` method in this class initialises its `c` field, which is a `Context` object for use in execution and solution verification, and calls the solution's `execute` method with `c` as an argument.

7. Checking that the problem is solved by the provided solution

Since different questions have different requirements, the method in `AbstractPExercise` that checks the validity of solutions, `checkSolved`, must also be abstract. This method returns a `boolean`, which will have the value `true` if the solution is correct and `false` otherwise. There are two possible ways of checking whether a solution is valid: either the value of an variable can be compared with a target value, or the result of a print statement can be inspected. The first of these approaches requires an `Id` field to be declared in the concrete class representing the problem, assigned the appropriate name in the constructor. In the overridden `checkSolved` method, the final value of this variable can be accessed by extracting the `value` field from the `ReturnValues` object attained by calling the `Id` instance's `evaluate` method with `c`, the `Context` object used during execution. In contrast, the second approach can be implemented

by checking the contents of the file written to by the representation of the print statement.

An example concrete class extending `AbstractPExercise` is `FactorialExercise`, which represents an exercise in which the student is asked to use a while loop to calculate the factorial of some integer N , which is supplied on construction of the object, storing the result in the variable `total`. The model solution code is shown below:

```
int total = 1, n = N;
while (n > 1) {
    total *= n--;
}
```

This code calculates $N \times (N-1) \times \dots \times 2 \times 1 = N!$ and stores it in `total` as required. In the `FactorialExercise` constructor, the value of $N!$ is calculated in the same way and stored in the `int` field, `nFact`, and the `totalID` field of type `Id` is set up with the name “`total`”. The overridden `checkSolved` method evaluates `totalID` with the `Context` object `c` and compares its final value with the `nFact` field, returning `true` if they are the same and `false` otherwise.

In this section I have presented the class `AbstractPExercise`, shown how it fulfils the requirements for representing an exercise, and shown a concrete example that extends it.

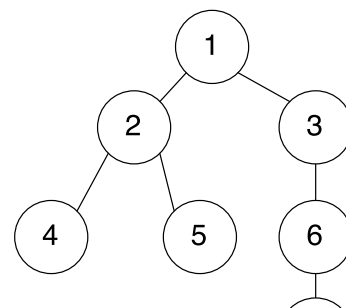
3.5 Adding and removing blanks

This section describes first the algorithm that adds a blank to a MiniJava program that can later be filled in by a student, followed by the algorithm that replaces a blank with the original code snippet.

When we try to add a blank to the model solution, it is possible that the entire solution is already blank, in which case we would fail. For this reason, the algorithms that add and remove blanks both return `booleans`, the value of which records if a blank was added or not.

The difficulty of a problem is based on the difficulty of the question itself and the percentage of the model solution that is blank when it is presented. Changing the question itself can be considered a coarse-grained adjustment to the problem’s overall difficulty, while changing the percentage of the model solution that is blank is a much more fine-grained adjustment of overall difficulty. When we wish to adjust the difficulty of the problem, we would prefer to have the greatest control over the difficulty possible, able to make the finest-grained adjustments. For this reason, the algorithm that chooses which node in a model solution to replace with a blank should make the choice that causes the smallest growth in the percentage of the whole that is blank. Additionally, we would prefer to distribute these blanks across the whole solution, rather than concentrating them in any one part.

Figure 3.2 shows an example AST, with its nodes numbered for reference. Using this tree we can compare different algorithms for selecting nodes, and see which best fulfils our requirements. Since we want to be able to add blanks to the



model solution until the entire solution is blank, we should consider traversal algorithms. The algorithm will terminate when a node is selected to be made blank, and when it is next run it should (effectively) pick up from the same point. Since, clearly, the first choice of node to be replaced should be a leaf (any other choice would cause a larger percentage increase of the number of blank nodes, since such a node would have children that would simultaneously be made blank), some standard algorithms suggest themselves. The Depth First Traversal algorithm, when run on the example tree, would produce the output: 4, 5, 2, 8, 7, 6, 3, 1. Notice that node 2 is visited before node 8. This algorithm would concentrate all the blanks in the left hand side of the tree, reaching up to the left-most child of the root before considering the right hand side of the tree. This renders the Depth First Traversal algorithm unfit for our purposes. Another potential traversal algorithm to consider is the Reverse Level Order Traversal algorithm, which acts something like a Breadth First Traversal from the leaves upwards. When run on the example tree, this algorithm would produce the output: 8, 7, 4, 5, 6, 2, 3, 1. This algorithm suffers from a similar problem: node 7 is visited before node 4. This results in the blanks being concentrated on areas of the tree that have the deepest nodes, neglecting all other areas. This again means that the Reverse Level Order Traversal algorithm is not fit for our purposes. The ideal output for our algorithm would be as follows: 4, 5, 8, 2, 7, 6, 3, 1. This does a good job of distributing blanks across the tree without reaching the upper levels of the tree too soon. Since the tree is unbalanced, once we have visited node 2 we have no choice but to visit the long chain of nodes on the right hand side of the tree, but at least there are a decent number of blanks on both sides of the tree at that point. This output can be attained by a traversal algorithm that operates as described below, but before that it is worth making a somewhat subtle clarification. When describing these algorithms it has been useful to think of them as traversal algorithms, even though I would not be using them that way. At each execution of the algorithm, a single node is made blank. On the next execution, a different node will be made blank. After a finite number of executions, the entire solution will be blank. Since every node is selected exactly once, the combination of all the executions leading to an empty model solution can be considered a traversal of the tree. The following algorithm is difficult to describe as a simple traversal, since that is not its aim. As such, the following description includes directions to remove nodes from the tree, something that you definitely wouldn't find in a standard traversal. The simplest description of this algorithm comes with the appreciation that the purpose of the algorithm is in fact to replace nodes with blanks³, not traverse the tree. With that in mind, here is a description of the ideal “traversal” algorithm:

1. Traverse the tree as in the Depth First Traversal algorithm.
2. Whenever a leaf is reached, return that node, and remove it from the tree.

³Replacing a node with a blank is similar enough to deleting the node entirely that they can be interchangeable in this context. This is explained in the following paragraphs.

3. Repeat until the root has been removed.

It can be seen that this algorithm does indeed produce the desired output, 4, 5, 8, 2, 7, 6, 3, 1. During the first pass of the tree, the leaves labeled 4, 5, and 8 are returned and removed. Removing these leaves means that nodes 2 and 7 are now the leaves of the tree. On the next pass through, these leaves are returned and removed. The “tree” is now a simple chain of the nodes (returned in the order of bottom to top) 6, 3, 1.

The problem my algorithm must solve is more complex than a simple traversal. Not only should my algorithm terminate mid-traversal and be able to pick up from that point at the start of the next execution, but the process of actually inserting a blank into the tree is non-trivial, and not only that, but the results of executing the algorithm have to be reversed in order to replace a blank with the original code snippet. To see how I implemented the algorithm, found in the `addBlank` method of `AbstractPExercise`, let us begin by considering the simple problem and introduce more complexity as we go.

Since the first step is to traverse the tree as a Depth First Traversal, we must first implement this, using a `Stack<MiniJASTNode>` to store the nodes that still need to be dealt with. We start by pushing the root node (found in the `solution` field of the `AbstractPExercise` class) on to the nodes stack, as usual, and as usual we process this node before adding its children to the stack. Point 2 of the algorithm description says that when we find a leaf, we remove it from the tree. In the actual algorithm we wish to replace the node with a blank (the process for which I explain shortly), but since we are do not actually remove the node from the tree, we have a problem. Once we have made all the leaves of the tree blank, we have no more leaves to choose; this is in direct contrast to when we were simply removing leaves thereby creating more leaves. To fix this problem, we redefine the concept of leaves: a node is a leaf if it has no children or all of its children are blank. In this way, by making leaves blank we are in fact creating more leaves. This definition of leaves is implemented in the `getIsLeaf` method declared in the `MiniJASTNode` interface and defined in the `ExpressionBase` and `StatementBase` abstract classes.

When we find a leaf, we wish to replace it with a blank. To do this, we have to replace the reference to the node in its parent with a reference to a new blank. At the moment, you cannot index nodes: you cannot access the n th child of a node, since they are simply stored in `private` fields. This means that it is very difficult to replace the reference to a child within a parent, without implementing a setter in every class that could translate the integer index to the appropriate field. This seems a lot more tiresome than simply allowing one to index nodes. To enable this, the `MiniJASTNode` interface declares the method `getSubNodes`, which returns `ArrayList<? extends MiniJASTNode>`. `StatementBase` and `ExpressionBase` both have a protected field, `subNodes`, which is returned by their overrides of this method. These fields are populated by the individual classes as their instances are set up. This means that rather than store their children directly as fields, each class in the implementation stores all its children in the `subNodes` field, and separately stores indexes into the array to provide access to the children during evaluation or execution. When we wish to replace a node with a blank, we store a blank at the appropriate index within its parent’s `subNodes` field using the `getSubNodes` method.

This means that we need to keep track of what the parent of the current node is, and also of the current node's index within its parents. These are stored in **Stacks** whose entries are updated as needed during the execution of the algorithm.

Since we need to be able to undo the effects of an execution of the algorithm to be able to restore the original code, we need to store the nodes that were replaced as well as the list of indices giving the path through the tree to each node's original location. These are stored as **private** fields of the **AbstractExercise** class, **Stack<MiniJASTNode> replacedNodes** and **Stack<Stack<Integer>> replacedNodeTreeIndices**. Whenever we replace a node, it is pushed onto the first stack, and the current stack of indices is pushed onto the second. The indices are also useful in enabling us to restart the algorithm where it left off. At the beginning of the algorithm, we check if the **replacedNodeTreeIndices** has any entries. If it does, we take the top one, reverse it (since the integer on top of the stack is the index of the last node in its parent and the integer on the bottom of the stack is the index of the root we need to follow first), and follow the indices through the tree. At each entry of the stack, we need to push the current node on to the parents stack, the index on to the index stack, and the children at and after, but not before, the index on to the nodes stack. The children after the index need to be pushed on too, as they have not yet been processed this pass of the tree.

In contrast to adding a blank, the algorithm that removes a blank from the model solution, **removeBlank** in **AbstractPEExercise**, is much simpler. We simply take the top off the **replacedNodes** and **replacedNodeTreeIndices** stacks, follow the indices to the parent of the specific node, and use the last index to reinsert the node back in to its parents **subNodes** field.

This section described the implementation of the algorithms that introduce blanks to and remove them from a MiniJava program.

3.6 The ExerciseSetter

In this section I describe how I made use of the language and its features to design the **ExerciseSetter**, one possible way in which students might interface with exercises designed to facilitate the learning of MiniJava.

The purpose of the **ExerciseSetter** class is to demonstrate one way in which students might interact with MiniJava and the question implementation described previously. It was designed with the possibility of being plugged in to yet more peripherals in mind, and as we will see in Section 3.7 the GUI makes use of it. The basic intuition behind its design comes from the imagined use case of students in a class where the student to teacher ratio is very high, and students may not have instant access to a teacher in the case that they find the presented exercises too easy or too difficult. The challenge in this scenario would be to programmatically detect that the student finds the problem too easy or too hard, and to present an easier or a harder problem respectively. The **ExerciseSetter** facilitates this by implementing a very simple performance heuristic and by storing several questions of various difficulties, such that ideally all students could be catered too. Using

the performance heuristic, the **ExerciseSetter** attempts to adjust the problems difficulty appropriately.

In order to programmatically present easier or harder questions, it is sensible to have the list of exercises sorted according to difficulty. To do this, **AbstractPExercise** actually implements the **Comparable** interface, comparing based on the difficulty of the questions involved. The list of exercises in **ExerciseSetter** is then kept sorted as new exercises are added.

The performance heuristic operates on the assumption that the performance of the students can be inferred from the number of attempts at solving the problem before their solution was correct and the comparison between the number of nodes in the model solution tree and the number of nodes in the student's solution tree. The more attempts a student required to solve the problem correctly, the more difficult they found the problem, it is assumed. Also, if a student used many more nodes to solve the problem than the model solution did, then it is likely that they haven't understood a core concept being tested: they may have unnecessarily unrolled a loop for example. This heuristic is found in the **reportPerformance** method of the **ExerciseSetter**, a method which returns a **float** value. If the returned value is positive, then the student supposedly did well, and the next problem should be harder, while if the returned value is negative the opposite is true. The magnitude of the float indicates the severity of the adjustment that should be made.

The class also provides an interface through which students may fill blanks by submitting their own code. There are two **fillBlank** methods in the **ExerciseSetter** class, one which takes a **MiniJASTNode** reference and uses the exercise's **fillBlank** method directly to fill the desired blank. The other takes a **String** argument containing the code with which to fill the blank. To get the representation of this code to actually fill the blank, of course the parser must be used, but this presents an interesting problem. Certain code snippets have the potential to be both **Expressions** and **Statements**. An example of this is the code snippet `total = 1`. This can be parsed as an assignment, and would thus be classed an **Expression**, or a variable declaration, where the surrounding context might be `int total = 1;`, which would categorise this snippet as a **Statement**. Fortunately the parser generated by the ANTLR tool has a solution to this, as the starting terminal symbol for the grammar is not fixed. The **ExerciseSetter** can check whether the blank being filled represents an **Expression** or a **Statement**, and then sets the entry point of the parser to be either **expression** or **blockStatement** appropriately.

If, once an exercise has been solved, the next exercise needs to be harder, according to the performance heuristic, then the **ExerciseSetter** should add blanks to the solution until either the exercise is hard enough, or the entire solution is blank. If the latter occurs, then the **ExerciseSetter** should attempt to present a harder problem and calculate how much of it should be blank to give the required difficulty. If instead the exercise needs to be easier, then the **ExerciseSetter** should remove blanks from the solution until either the exercise is easy enough, or there are no blanks in the solution. If the latter occurs, then the **ExerciseSetter** should attempt to present an easier problem and calculate how much of it should be blank. This functionality is implemented in the **adjustQuestion** method of the **ExerciseSetter**.

To enable it to be plugged in to other peripherals, the **ExerciseSetter** simply propagates any custom exceptions raised during execution of the solution upwards, and returns **Strings** when asked to present the current question or print the current solution.

In this section I have detailed one of the ways that students might interface with MiniJava through the **ExerciseSetter**. This is only an example of a possible use for the tools, but shows how powerful they can be.

3.7 The GUI

In this section I describe the implementation and function of the GUI peripheral.

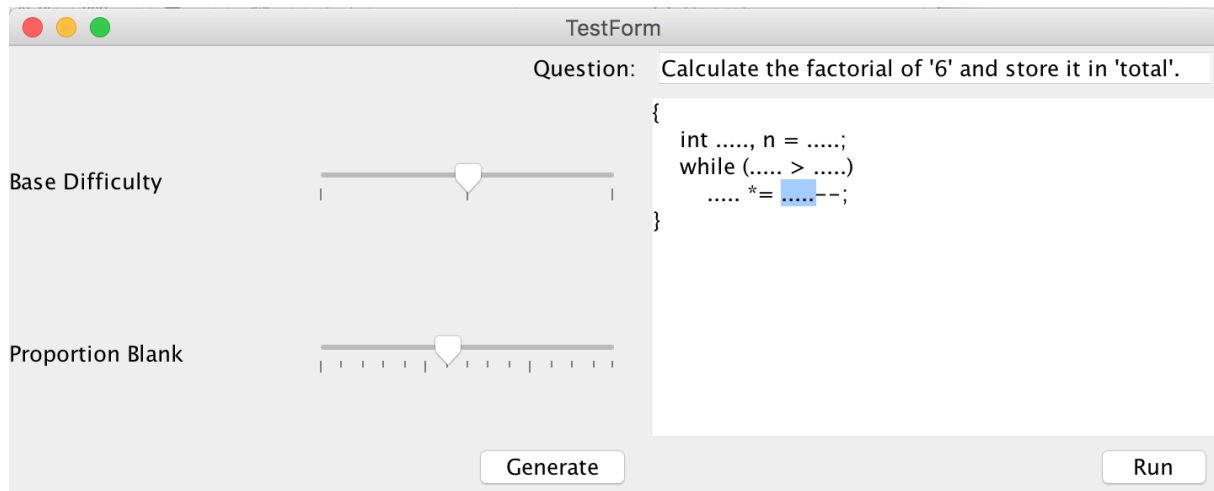


Figure 3.3: An example screenshot taken of the GUI peripheral.

The GUI demonstrates some of the possible uses of this set of tools. It shows how a teacher might manually set the difficulty of a particular exercise and how a student could fill in the blanks and run their solution. An example screen from the GUI is shown in Fig 3.3. The top slider determines the difficulty of the question: the further to the left the slider is, the easier the question will be. The bottom slider affects the difficulty of the problem: the further it is to the left, the fewer blanks there will be and thus the easier the problem will be to solve. When the “Generate” button is clicked on, the corresponding exercise will be displayed in the box on the right.

As shown in the screenshot, the box on the right is editable, allowing students to directly input their own code, filling in the blanks or otherwise. When the “Run” button is clicked on, the code in this box is parsed and set as the solution to the exercise. This solution is then run, and the validity of the solution checked. The GUI reports whether the solution was correct or not, and attempts to present error messages in a user friendly manner if any were produced.

The GUI was designed using the IntelliJ UI Designer plugin⁴. This plugin allowed me to drag and drop components into place on the form before programming their functionality with the adjoining bound class. The class stores a reference to an **ExerciseSetter**

⁴<https://www.jetbrains.com/help/idea/2017.1/swing-designing-gui.html>

object, which is where all the data comes from, and where the commands from the user are delivered to.

This section described the function and then the implementation of the GUI peripheral.

In this chapter I have described what I have implemented and how I have done it. I started by giving the design of the language MiniJava, explaining some of these choices by introducing the parser and the concept of “blanks” within a MiniJava program. From there I explained how a programming exercise is conceptualised in the abstract sense, and gave an example of an implementation of a specific question. I finished by giving two possible uses for the set of tools, the **ExerciseSetter** and the Graphical User Interface (GUI).

Chapter 4

Evaluation

This chapter demonstrates that all of the original goals of the project were achieved. In Section 4.1 it is shown that all the components fulfil the requirements laid out earlier. The trends that emerge as the difficulty is adjusted are analysed in Section 4.2, and the development process is critiqued in Section 4.3.

4.1 Correct functionality

A test suite consisting of a total of 113 tests (all of which pass when run on the current source code) demonstrates that the aims of the project are satisfied by my implementation. This section demonstrates the correct functionality of some of the more complex components.

4.1.1 The `ExerciseSetter` class

The aim of the `ExerciseSetter` class is to be capable of delivering programming exercises, measuring the performance of students who solve these exercises, and adapting the difficulty of the following exercises appropriately. It has been shown in Section 3.6 that the `ExerciseSetter` class can deliver exercises by printing the question and a `String` representation of the incomplete model solution. In Section 3.5 the algorithm that adapted the difficulty of exercises was shown. All that remains is to analyse the way performance is measure.

The class has a method `reportPerformance` which returns a `float` value that represents the performance of a student once they have solved an exercise. If this value is positive, the student has performed well enough to be given a more difficult exercise, whereas if it is negative then the next exercise will be easier. Figure 4.1 shows how the measured performance depends on both the number of attempts a student has made to give a correct solution and also the number of extra nodes they used in their solution compared with that of the model solution. This graph shows the contours of the surface, which is a function of two variables. The orange line labelled 0 represents the zero contour, such that any solution that maps to a point above that line (on the 2D plot) is unsatisfactory, and will result in the next exercise being easier. Solutions mapping to points below the zero contour are good enough that the next exercise will be harder.

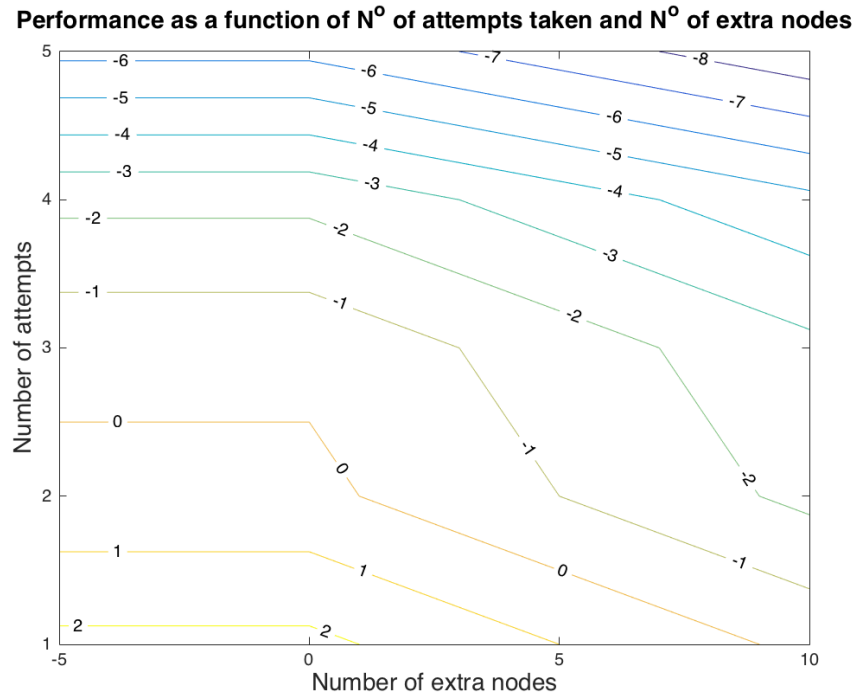


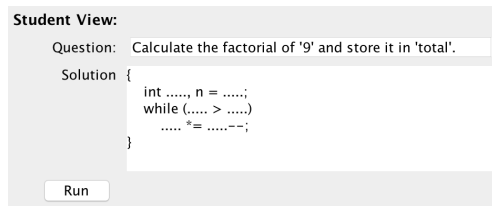
Figure 4.1: A graph showing the contours of the performance function.

The performance varies as a function of the number of attempts squared, strictly punishing solutions that required many attempts, as this indicates a lack of understanding of the question being answered. The cut-off point for the number of attempts is 2.5, that is if a student requires two attempts or fewer, the system determines that the next exercise can be harder, but if a student requires three attempts or more, the next exercise will be easier. This was chosen to allow one careless mistake before assuming lack of comprehension.

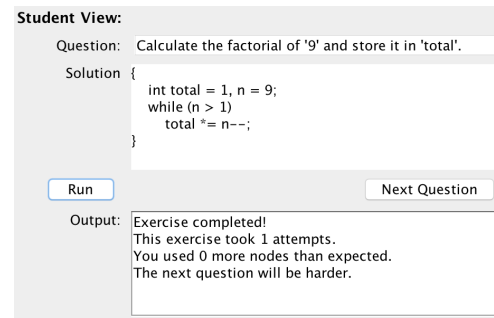
If the student uses fewer nodes than the model solution, then either they have found a more efficient solution, or they have “cheated” and not used the appropriate syntax constructs being taught (i.e. reached $n!$ by simply calculating $n \times (n - 1) \times \dots$ all at once rather than using the expected while loop). At present the system is not able to tell the difference, but an area of future work might be to enable the system to check for certain code constructs being used in the presented solution. Since, at the moment, a student using fewer nodes could be indicative of understanding or of cheating, the performance function does not take this into account, shown by the horizontal contours to the left of the $x = 0$ vertical. When the number of extra nodes used is greater than zero, it is almost certainly a sign that a student has not understood the best way to solve the question (i.e. unrolled the while loop to calculate $n!$). For this reason, the greater the number of extra nodes used, the worse the measured performance.

4.1.2 The parser extension

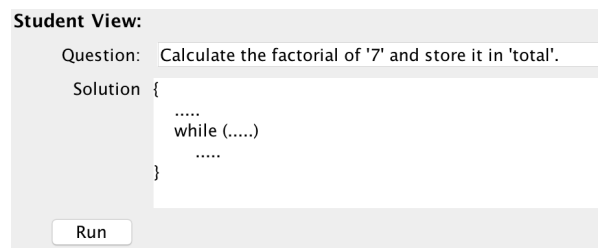
The functionality of the parser is ensured by the continued running of regression tests after every change is made. Some of these tests are in the test suite itself, but the parser is also



(a) The initial question presented to the student.



(b) The result of solving the question in one attempt and with no extra nodes.



(c) The harder question resulting from this good performance.

Figure 4.2: Screenshots of the GUI when the student has performed well.

used to simplify the usage of some of the other components. For example, the exercises stored in the `ExerciseSetter` store their model solution as a `String` that is then parsed to allow the solution to be executed. Thus, whenever an exercise is instantiated, the parser is also being regression tested.

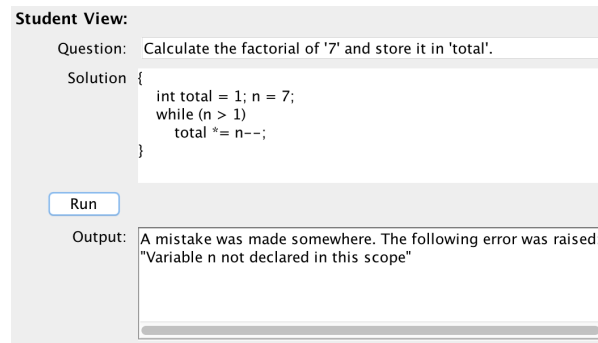
Since the GUI also makes use of the parser to allow students to directly edit the model solution however they choose, any use and evaluation of the GUI also functions as testing of the parser.

4.1.3 The GUI extension

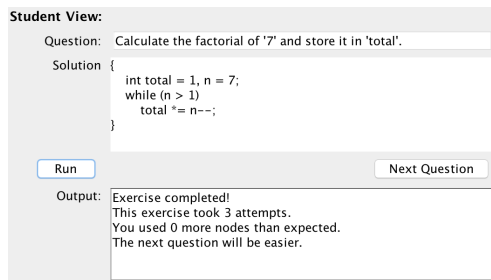
Investigating the GUI is the best way to evaluate the system as a whole, since it brings together all of the components and provides an interface for users. The screenshots in Figure 4.2 show the output given when a student solves the exercise in the best way possible, in one attempt and with no more nodes than expected.

The result of solving the question perfectly (shown in Figure 4.2b) is recorded in the “Output:” part of the GUI, which gives the statistics of the solution (i.e. then number of attempts and the number of nodes used) and states that the next question will be harder. This question (shown in Figure 4.2c) is presented when the student presses the “Next Question” button. While a solution has only one way of being successful, there are a variety of ways in which it can be unsatisfactory. Some possible ways are shown in Figure 4.3.

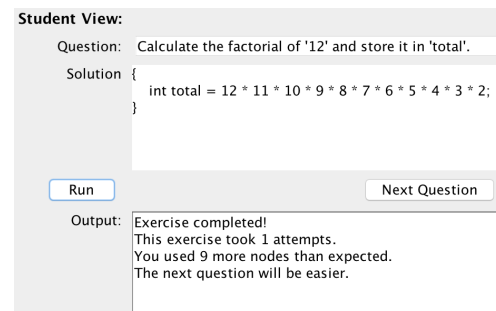
When a mistake is made, the error thrown by the interpreter is propagated up and shown by the GUI, as demonstrated by Figure 4.3a. This should hopefully help the student



(a) A mistake has been made .



(b) The result of solving the question in one attempt and with no extra nodes.



(c) The harder question resulting from this good performance.

Figure 4.3: Screenshots of the GUI when the student has performed well.

to find the error they made and correct it. In this case, the error tells us that “The variable `n` has not been declared”. Indeed, in the first line of the code, the punctuation after the word “total” should be a comma, not a semi-colon, to show that “`n = 7`” was part of the local variable declaration, not an assignment. For this particular solution, we imagine that the student made another mistake (which is not shown), before finally submitting a correct answer as in Figure 4.3b. Since the student took three attempts to solve the exercise, the system determines that the student has not properly understood the exercise so, as stated in the “Output:” box, the next exercise will be easier. Another possible issue a student could have is not understanding how to use the appropriate code construct to attain the most efficient result. This is shown in Figure 4.3c, where rather than use a while loop to calculate a factorial they have simply written out the whole long calculation. It is debatable which of the two solutions is preferable in general, but given the purpose of this exercise was to teach how to use while loops, I decided that the given solution was unsatisfactory. This can be seen in the “Output:” panel that lists that even though the problem was solved in one attempt, nine more nodes were used than expected, so the next question will be easier.

4.2 Trends in difficulty against performance

In Section 4.1.1 I analysed the way that the `ExerciseSetter` class measured the performance of a student through their submission. In this section I analyse how this measure-

Difficulty increase when repeated submissions have a performance of 1

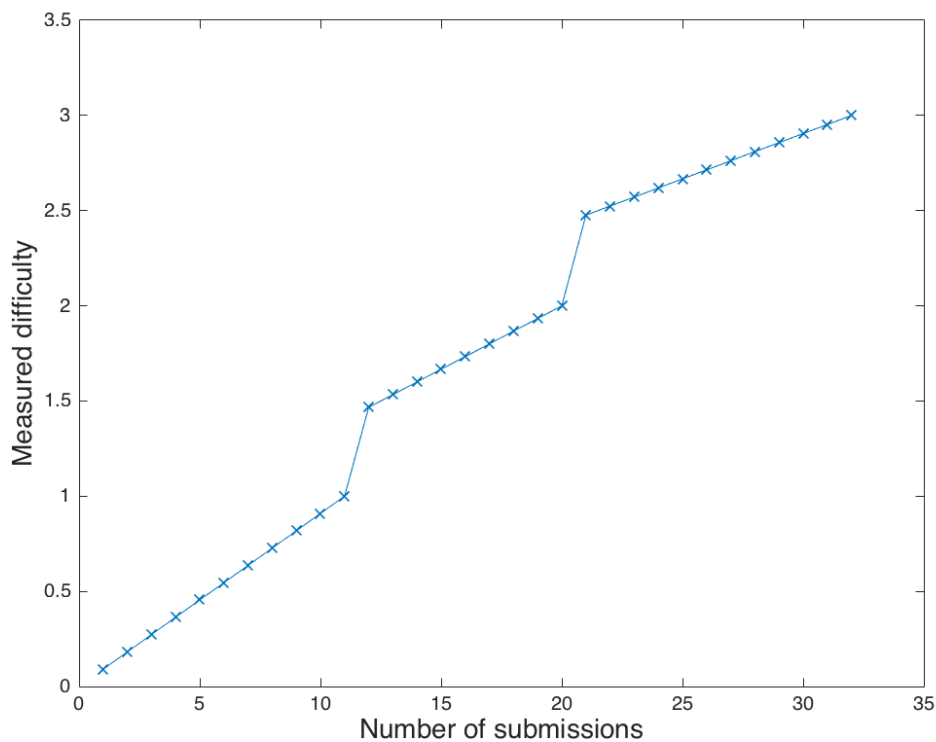


Figure 4.4: Increase in measured difficulty with repeated submissions each with a measured performance of one.

ment is used to affect the difficulty of following exercises, and the trends that arise from this method.

Figure 4.4 shows how the difficulty of exercises presented by the `ExerciseSetter` increases over a number of submissions. To generate this graph, I set the difficulty of the exercise being delivered by the `ExerciseSetter` to be as low as possible, that is, the simplest question with only a single blank in its model solution. I then made repeated correct submissions, tailored such that they all had a measured performance of one. In practice this is unrealistic, as it corresponds to the understanding of the student increasing at exactly the same rate as the difficulty, however, analysing this result can still be instructive. The difficulty of the exercise was increased by adding blanks to the model solution until the entire solution was blank. After the next correct submission, the next most difficult question was set.

At present, there are three questions that the `ExerciseSetter` can ask. These can be seen in the graph as the three distinct groups of points, separated by quite severe increases in difficulty. These sudden increases can be explained by my choice of how the system should adjust the difficulty when a new question is asked for the first time. Instead of giving the exercise the lowest difficulty possible (i.e. leaving only a single blank in its model solution), the system sets half of the model solution to be blank. This was done to ensure that a student trying a new problem for the first time would not be given the majority of the code at first attempt.

For the three example questions I have implemented to demonstrate the power of

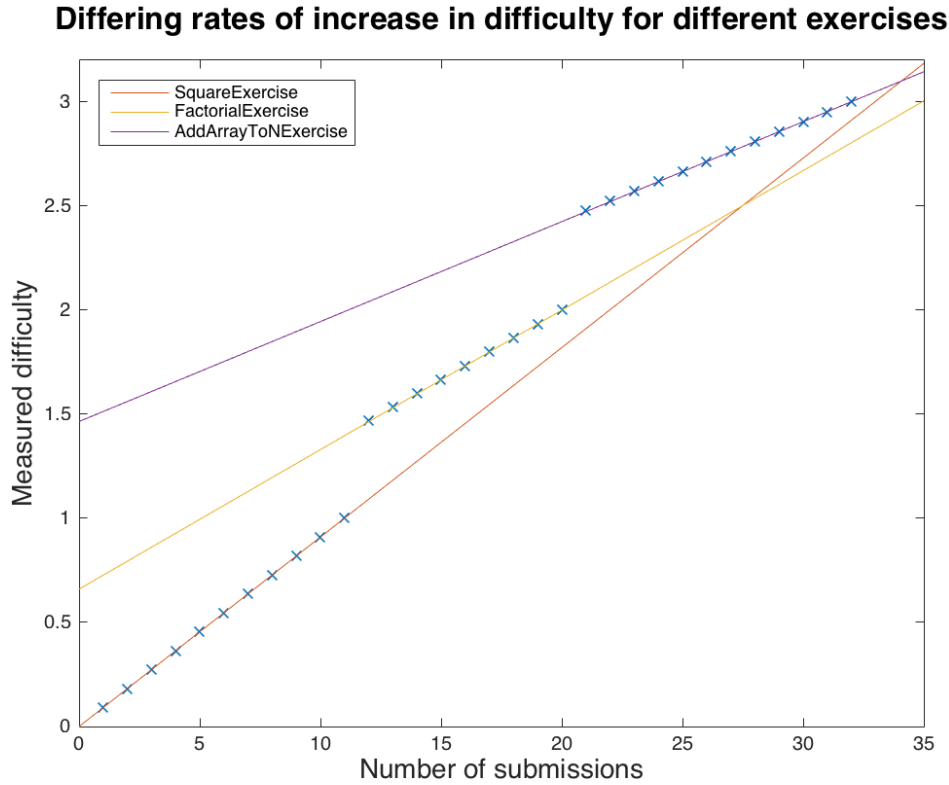


Figure 4.5: This demonstrates the different rates of increase of difficulty found for the different questions. The straight lines are extrapolated past the original points to show that they intersect and thus are not parallel.

these tools, it is a trend that the more complex the question, the more nodes the model solution requires to solve the problem. This has the unintended effect of making the rate of increase in difficulty slower as the difficulty of the question itself increases. This trend can be seen more explicitly in Figure 4.5, where each solid line corresponds to a single question. Since the gradient of these lines decreases as the questions get more difficult, so too does the rate of increase in difficulty across submissions. This is a trend that we might desire in our system; it might be expected that students will need more support as the questions get more difficult, so it would be useful for the rate of increase in difficulty to be slower for harder questions.

Although this trend may not hold for all questions (i.e. there may be some simple questions that require a lot of nodes to answer, and some complex ones that can be solved with very few nodes), I believe that it will hold for a large number of questions, making this a useful side effect of my particular implementation.

In this section I have described the trends present in the way difficulty is modified according to a student's performance.

4.3 Development process

Throughout the implementation process, I followed the principles of Test Driven Development (TDD)¹. This methodology revolves around the writing of unit tests before the appropriate code is written. These tests are designed with reference to the requirements specification, such that any software that passes these tests fulfils the requirements. I found it very useful to develop the software in this manner, and would want to incorporate this methodology in my next large project.

The creation of prototypes to further investigate certain ideas has also been beneficial, allowing me to avoid expending too much effort on a method that would not have worked. However, at times I might have benefited from strictly laying out the interfaces between the components before I started to make prototypes of them. Sometimes the prototypes became the components themselves, and it was only when I tried to get them to work together that I realised that the interface one component expected was not necessarily the one that another component offered. Starting again now, I would outline the project not only in text form, but also as a UML diagram.

This section has evaluated the functionality of the individual components, some of the trends found in the data, and my own development process.

¹<http://agiledata.org/essays/tdd.html> - last accessed 10th May

Chapter 5

Conclusion

In conclusion, the project was a success. The software I produced fulfils all of the original aims, and I was able to implement two extensions on top of the core functionality that have enabled me to properly demonstrate the power of the underlying system.

I have designed a new language, called MiniJava, such that its programs can incorporate gaps that must be filled in so that they can be executed. I have written an interpreter and a pretty-printer for this language, to allow arbitrary code snippets to be executed, and so that internal representations of these programs to be printed using the concrete semantics. I encoded a representation of programming exercises to be solved using MiniJava, and wrote algorithms to introduce blanks to and remove them from the model solutions for these exercises, making them harder and easier respectively. Using concrete examples of these exercises, I implemented a system that could deliver questions to a student, allow them to fill in the blanks with their own arbitrary code, and execute the complete solution to determine whether the problem was solved. This system measures the performance of the student using a heuristic, and adjusts the difficulty of the next exercise based on this measurement. As extensions, I implemented a parser for MiniJava to streamline the process of accepting student input, and created a GUI to demonstrate how both teachers and students might use the system.

This project was subject to several limitations, not least of which was my lack of expertise in the field of teaching. A lot of the choices I made were based on my intuition, and given the time and resources I would have benefitted greatly from the guidance of an education professional throughout the development process. Further to this, the system suffers from a paucity of possible questions that may be delivered, due to a lack of time and, again, expertise. Given more variety of questions of a similar difficulty, the system would not need to present the same question multiple times, as currently it is very possible for students to simply learn the answers without understanding the material.

5.1 Further work

Part of writing systems that attempt to teach programming that is often very difficult to manage is the reporting of errors in a way that novice programmers would find useful. One tool that makes some headway in solving this problem is the Arjen tool[4]. This

tool exploits the fact that novice programmers all tend to make the same types of errors, and explicitly searches for these errors in order to provide feedback tailored to a novice, perhaps providing pointers for how to solve or avoid the problem. A collaboration between my set of tools and a tool such as Arjen would certainly be beneficial for students, giving a much more supportive user experience.

Another extension I considered in my proposal but did not have time for in my implementation was to consider different syntactic constructs separately, and attempt to record a student's expertise with each construct. This would allow a much more thorough analysis of a student's understanding, and allow the system to provide a more tailored and directed learning experience.

Bibliography

- [1] Jayoung J Goo, Kyoung S Park, Moonhoen Lee, Jieun Park, Minsoo Hahn, Hyungil Ahn, and Rosalind W Picard. Effects of guided and unguided style learning on user attention in a virtual environment. In *International Conference on Technologies for E-Learning and Digital Entertainment*, pages 1208–1222. Springer, 2006.
- [2] J. Gosling, B. Joy, G.L. Steele, G. Bracha, and A. Buckley. *The Java Language Specification*. Java (Addison-Wesley). Addison Wesley Professional, 2014.
- [3] Terence Parr. *The definitive ANTLR 4 reference*. Pragmatic Bookshelf, 2013.
- [4] Warren Toomey. Quantifying the incidence of novice programmers errors. *School of IT, Bond University*.
- [5] Diana F Wood. Problem based learning. *BMJ: British Medical Journal*, 326(7384):328, 2003.
- [6] Chye Foong Yong. *Automatically generating adaptive programming exercises based on student profiles*. University of Nottingham, 2007.

Appendix A

The Parser

```
grammar MiniJava;

// STATEMENTS / BLOCKS

// entry point
entry
    : block [true]          # blockEntry
    | blockStatement+       # blockStatementsEntry
    | statementTop          # statementEntry
    | expression            # expressionEntry
    ;

block [boolean isOuter]
    : LBRACE blockStatement* RBRACE
    ;

blockStatement
    : primitiveType variableDeclarators SEMI    # localVariableDeclaration
    | statement                                # makeStmnt
    | variableDeclarator                        # makeVarDec
    ;

statementTop
    : statement                # stmnt
    | statementNSI             # stmntNSI
    ;
```

```

statement
    : block [false]                                # makeBlock
    | IF parExpression statement                    # makeIf
    | IF parExpression statementNSI ELSE statement # makeITE
    | FOR LPAREN forInit? SEMI expression?
      SEMI expressionList? RPAREN statement        # makeFor
    | WHILE parExpression statement                # makeWhile
    | statementNTS                                  # makeStatementNTS
    ;

statementNSI
    : block [false]                                # makeBlockNSI
    | IF parExpression statementNSI ELSE statementNSI # makeITENSI
    | FOR LPAREN forInit? SEMI expression?
      SEMI expressionList? RPAREN statementNSI      # makeForNSI
    | WHILE parExpression statementNSI              # makeWhileNSI
    | statementNTS                                  # makeStatementNTSNSI
    ;

statementNTS
    : DO statement WHILE parExpression SEMI        # makeDo
    | RETURN SEMI                                  # return
    | BREAK SEMI                                   # break
    | CONTINUE SEMI                               # continue
    | SEMI                                          # empty
    | expressionStatement SEMI                    # makeStmntExpr
    ;

forInit
    : primitiveType variableDeclarators            # forInitLVD
    | expressionList                               # forInitExprs
    ;

// EXPRESSIONS

parExpression
    : LPAREN expression RPAREN
    ;

expressionList
    : expression (COMMA expression)*
    ;

```

```

expressionStatement
: expression
;

expression // Most binding comes first!
: Identifier # makeID
| expression LBRACK expression RBRACK # arrayAccess
| parExpression # makeBracketed
| literal # makeLiteral
| expression (op=INC | op=DEC) # postInc
| (op=ADD|op=SUB|op=INC|op=DEC) expression # preIncEtc
| BANG expression # makeNot
| expression (op=MUL|op=DIV) expression # multExpr
| expression (op=ADD|op=SUB) expression # addExpr
| expression (op=LE | op=GE | op=GT | op=LT) expression # relationalExpr
| expression (op=EQUAL | op=NOTEQUAL) expression # eqExpr
| expression AND expression # andExpr
| expression OR expression # orExpr
| <assoc=right> expression QUESTION
    expression COLON expression # condExpr
| <assoc=right> expression
    ( op=ASSIGN
      | op=ADD_ASSIGN
      | op=SUB_ASSIGN
      | op=MUL_ASSIGN
      | op=DIV_ASSIGN
    )
    expression # assignExpr
;

// VARIABLES AND LITERALS

variableDeclarators
: variableDeclarator (COMMA variableDeclarator)*
;

variableDeclarator
: Identifier LBRACK RBRACK (ASSIGN variableInitializer)? # arrayVarDec
| Identifier (ASSIGN variableInitializer)? # singleVarDec
;

```

```

variableInitializer
    : arrayInitializerValues                # arrayInitVals
    | arrayInitializerSize                  # arrayInitSize
    | expression                            # initExpr
    ;

arrayInitializerValues
    : LBRACE variableInitializer (COMMA variableInitializer)* (COMMA)? RBRACE
    ;

arrayInitializerSize
    : NEW primitiveType LBRACK expression RBRACK
    ;

primitiveType
    : BOOLEAN
    | CHAR
    | INT
    | DOUBLE
    ;

literal
    : IntegerLiteral
    | FloatingPointLiteral
    | CharacterLiteral
    | BooleanLiteral
    ;

// LEXER
// 3.9 Keywords

BOOLEAN      : 'boolean';
BREAK        : 'break';
CHAR         : 'char';
CONTINUE     : 'continue';
DO           : 'do';
DOUBLE       : 'double';
ELSE         : 'else';
FOR          : 'for';
IF           : 'if';
INT          : 'int';
NEW          : 'new';
RETURN       : 'return';
WHILE        : 'while';

```



```
// 3.10.1 Integer Literals
// 3.10.2 Floating-Point Literals
// 3.10.3 Boolean Literals
// 3.10.4 Character Literals
// 3.11 Separators
// Sections removed for clarity
```

```
LPAREN      : '(';
RPAREN      : ')';
LBRACE      : '{';
RBRACE      : '}';
LBRACK      : '[';
RBRACK      : ']';
SEMI        : ';';
COMMA       : ',';
DOT         : '.';
```

```
// 3.12 Operators
```

```
ASSIGN      : '=';
GT          : '>';
LT          : '<';
BANG        : '!';
QUESTION    : '?';
COLON       : ':';
EQUAL       : '==';
LE          : '<=';
GE          : '>=';
NOTEQUAL    : '!=';
AND         : '&&';
OR          : '||';
INC         : '++';
DEC         : '--';
ADD         : '+';
SUB         : '-';
MUL         : '*';
DIV         : '/';

ADD_ASSIGN  : '+=';
SUB_ASSIGN  : '-=';
MUL_ASSIGN  : '*=';
DIV_ASSIGN  : '/=';
```

```
// 3.8 Identifiers (must appear after all keywords in the grammar)
//
// Whitespace and comments
//
// Sections removed for clarity
```

Appendix B

Project Proposal

Computer Science Tripos – Part II – Project Proposal

A System for Giving Programming Exercises Adaptive Difficulty

R. J. McFarland, Homerton College

Originator: M. B. Gale

20 October 2016

Project Supervisor: M. B. Gale

Director of Studies: Dr J. Fawcett

Project Overseers: Dr D. J. Greaves & Prof J. G. Daugman

Introduction

Different people learn at different speeds, and learn some things more quickly than others. When a group of students are given a set of programming exercises, this typically isn't taken account of. The aim of this project is to design a system that varies the difficulty of a programming exercise depending on how the individual has performed completing previous exercises.

The difficulty of an exercise can be measured using two metrics: how complex the problem being solved is and how much code the student is expected to write. The system will vary both of these to adapt the difficulty of proposed exercises automatically (i.e. the system will not look for preprogrammed mistakes, but rather interpret the mistakes made and react accordingly). When an exercise is completed to a sufficient standard, a similar exercise will be given, but expecting more code to be written. When the system has determined that the problem has been mastered, a more difficult problem will be presented, requiring less code to be written again.

Resources required

I shall use my own Macintosh laptop for the majority of this project. Backup will be to GitHub and to a 5TB hard drive I keep in my room. I shall make use of the Java standard library, as well as the JavaFX library for the graphical user interface element. I require no special resources.

Starting point

I am able to program in Java, and have used JavaFX before for the Part IB group project. During my A-Levels I wrote a program to test the maths abilities of Year 4 students, which involved the programmatic generation of various kinds of maths questions.

Work to be done

The work for this project can be split up into the following sub-projects:

1. A representation of a programming exercise must be coded up to allow the final product to manipulate them.
2. A heuristic that measures the difficulty of a given exercise must be chosen and implemented.
3. A heuristic that assesses how well a “student” has completed an exercise must be chosen and implemented.
4. I must devise a system, using the previous three items, that determines which exercise should be given to the student next, as well as how much of the required code is already filled in, based on the student’s performance completing previous exercises.
5. A Graphical User Interface should be designed to enable a user to manually adjust the content of a given programming exercise.
6. In order to test the program, a student will be simulated by creating a system that tells the program what errors were made where and how long the exercise took to be solved, so that the program’s response to various stimuli can be measured.

Success Criteria

The project will be considered completed when:

1. I have a system that presents the same initial problem to all students.
2. The system can determine that code written by the student solves the given problem.

3. On registering completion of the exercise, the system will then present a new problem to the student.
4. If the student solved this problem with sufficient ease as measured by my performance heuristic, then this new problem will be more difficult, as measured by my difficulty heuristic.
5. If the student failed to solve the problem, or took too much time or made too many errors, then the next problem will be easier.
6. The system will adjust the difficulty of a given problem in line with the difficulty heuristic by changing the amount of code required to solve the problem, and also the underlying problem being solved.

Possible extensions

- This system could be improved such that it identifies which concepts a student is struggling with and adjusts the content of future exercises to encourage learning of these concepts.
- The program could produce a graph of progress against time to motivate learners.
- A simple game that utilises this system could be made, to showcase how the system might be used to teach programming.

Timetable

24th Oct - 6th Nov:

- Research adaptive difficulty and see how it has been done before.
- Research how errors in code have been quantified in the past.
- Research where and how simulated users have been designed to test systems.

7th Nov - 20th Nov:

- Describe how code that solves an exercise could be split up into discrete sections, and a way to store how well each of those sections was completed.

Milestones:

- Have a representation of a programming puzzle in Java code.

21st Nov - 4th Dec:

- Prototype and compare heuristics to measure the “difficulty” of a given programming exercise, based on the difficulty of the goal to be achieved, and how much of the code is presented initially.

Milestones:

- Select and implement a difficulty heuristic.

5th Dec - 18th Dec:

- Prototype and compare heuristics to measure performance, based perhaps on how much time was taken, the number of errors made and the time the solution takes to run.

Milestones:

- Select and implement a performance heuristic.

19th Dec - 1st Jan:

- Slack time over Christmas to catch up if I need to.

2nd Jan - 15th Jan:

- Implement a system to adapt the difficulty of the programming exercises by presenting a new, more difficult problem each time an exercise is completed satisfactorily, and presenting an easier one if the student is struggling too much.

Milestones:

- Have a system that will present a problem to a student, determine that the student has solved the problem with the code they have written, determine if the student found it too easy or too hard, and present the student with a new problem that is either easier or harder respectively.

16th Jan - 29th Jan:

- Write Progress Report.
- Improve the system by implementing the concept of changing the amount of code that needs to be written to solve a given problem as another way of affecting its difficulty.

Milestones:

- Have the system extended such that it may require the student to write more or less code to make more fine grained adjustments to the difficulty of a programming problem.

30th Jan - 19th Feb:

- **3rd Feb:** Submit the Progress Report
- Design a user interface to change the content of the programming exercise to be solved.

Milestones:

- Have a user interface with controls to determine the exact difficulty of the problem about to be presented.

20th Feb - 5th Mar:

- Write unit tests for all units.

Milestones:

- Have unit tests written for every unit of the code.

6th Mar - 19th Mar:

- Test the implementation by simulating a student “using” the system by “solving” problems with varying success.

Milestones:

- Have a simulated student that will tell the system what mistakes were made in the code and where, as well as how long it took to solve the problem (along with any other parameters deemed important to measuring the performance of the student), such that the response of my system can be measured against the performance of a student using it.

20th Mar - 2nd Apr:

- Slack time for catching up and doing extensions.

3rd Apr - 16th Apr:

- Write Introduction and Conclusion (about 2,500 words).
- Write up Proforma (excluding word count), Declaration of Originality, Project Proposal and Cover Sheet.

Milestones:

- Have the above sections of the dissertation written in draft form.

17th Apr - 30th Apr

- Write Preparation and Implementation (about 5000 words).

Milestones:

- Have the above sections of the dissertation written in draft form.

1st May - 14th May

- Write Evaluation (about 2500 words).
- Write Contents Page, finish Bibliography, Appendices and Index as required.

- Fill in word count of Proforma.
- Submit draft to DoS and Supervisor.

Milestones:

- Have the above sections of the dissertation written in draft form.

15th May - 19th May

- Incorporate feedback into dissertation.
- **19th May:** Submit dissertation.