VDMJ Design Specification

Author	Nick Battle
Date	08/04/09
Issue	0.4

0. Document Control

0.1. Table of Contents

0.2. References

[1] Wikipedia entry for The Vienna Development Method, http://en.wikipedia.org/wiki/Vienna_Development_Method

[2] Wikipedia entry for Specification Languages, http://en.wikipedia.org/wiki/Specification_language

[3] The VDM Portal, http://www.vdmportal.org/twiki/bin/view

[4] The VDMTools VDM-SL Language Manual, http://www.vdmtools.jp/uploads/manuals/langmansl_a4E.pdf

[5] The VDMTools VDM++ Language Manual, http://www.vdmtools.jp/uploads/manuals/langmanpp_a4E.pdf

[6] DBGP - A common debugger protocol for languages and debugger UI communication, http://xdebug.org/docs-dbgp.php.

[7] Overture - Open-source Tools for Formal Modelling, http://www.overturetool.org/.

0.3. Document History

Issue 0.1	22/10/08	First release.
Issue 0.2	28/11/08	Added comments from PGL. Added AST converter.
Issue 0.3	04/03/09	Added PO generator section, vdm.traces and misc other changes.
Issue 0.4	08/04/09	Added the DBGp protocol section.

1. Overview

VDMJ provides tool support for the VDM-SL and VDM++ specification languages, written in Java. The tool includes a parser, a type checker, an interpreter, a proof obligation generator and a debugger. It is a command line tool only, though it is accessible from graphical environments like Eclipse.

1.1. Package Overview

The implementation is divided into 19 Java packages, which are all sub-packages of org.overturetool.

Packages	
vdmj	The main VDMJ class and supporting classes.
vdmj.lex	Classes that implement the lexical analyser and its tokens.
vdmj.syntax	Classes that implement the syntax analyser.
vdmj.ast	Classes that translate an Overture AST to VDMJ's internal tree.
vdmj.types	Classes that represent static types during type checking.
vdmj.expressions	Classes that represent VDM expressions.
vdmj.statements	Classes that represent operation statements.
vdmj.patterns	Classes that represent patterns and binds.
vdmj.traces	Classes that represent trace definitions.
vdmj.definitions	Classes that represent VDM definitions.
vdmj.modules	Classes that represent VDM-SL modules and their import/export definitions.
vdmj.typechecker	Classes that support the static type checker.
vdm.pog	Classes that support the proof obligation generator.
vdmj.runtime	Classes that implement the interpreter.
vdmj.values	Classes that represent runtime values in the interpreter.
vdmj.commands	Classes that read and execute commands from standard input.
vdmj.messages	Classes that hold VDMJ error and warning messages.
vdmj.debug	Classes that implement the DBGp protocol.
vdmj.util	Utility classes and library routines used by all other packages.

The vdmj package contains the “main” abstract class for the suite, with two subclasses to parse, type check and interpret a specification in the VDM-SL or VDM++ dialect.

The vdmj.lex package contains all classes to do with the lexical analysis of specifications. This includes the lexical token reader, plus a set of value classes for representing the various types of lexical token.

The vdmj.syntax package contains all the syntax analysis classes. This includes eight “readers”, which implement a recursive descent parser, based on a stream of lexical tokens. The readers are all sub-classes of an abstract Reader class.

The vdmj.ast package contains classes to translate an Overture parsed AST into VDMJ's internal tree format. This is to permit the Overture parser to be used by VDMJ, but without changing the type checking and runtime classes.

The vdmj.types package contains value classes that represent the various static types that can be contained in a VDM specification. They are all sub-classes of an abstract Type class.

The vdmj.expressions package contains value classes that represent the various types of expression that can be defined in a specification. They are all sub-classes of an abstract Expression class, which defines methods for an expression to be type checked and evaluated. Similarly, the vdmj.statements package defines a set of value classes that subclass Statement and represent the different statements in a specification.

The vdmj.patterns package includes value classes to represent the various patterns and binds in a specification. They are all subclasses of an abstract Pattern or Bind class.

The vdmj.traces package includes classes that represent the possible trace definitions that can be declared in a VDM++ class.

The vdmj.definitions package contains a set of classes representing the definitions in a specification. They are all sub-classes of an abstract Definition class. Definitions are nested, so for example a class definition may contain function definitions, which in turn may contain function definitions for their pre and post condition functions. All definitions implement methods to perform type checking, and to generate runtime values representing their content.

The vdmj.modules package contains value classes that represent the modular structure of VDM-SL specifications, including their import and export declarations.

The vdmj.typechecker package includes classes which control the type checking of VDM-SL and VDM++ specifications. Most of the type checking is performed by the definition, expression and statement classes that collectively describe the specification, but the typechecker package defines the supporting classes to invoke the type checking methods of the other objects, and to represent the static environment in which type checking is performed.

The vdmj.pog package contains classes that support the proof obligation generator. Like type checking, the actual process of proof obligation generation is performed by the definitions, expressions and statements which form the specification, but the pog package contains classes to represent proof obligations.

The vdmj.runtime package defines the abstract interpreter, and its subclasses to interpret VDM-SL and VDM++ specifications. It includes classes to represent the runtime execution context, exceptions which can be generated at runtime, and the debugging and thread control classes.

The vdmj.values class contains a set of classes to represent runtime values in the interpretation of a specification. They are all subclasses of the abstract Value class. All values are immutable, with the exception of the UpdatableValue class, which has a set method and is used to implement all state variables in the system.

The vdmj.commands package contains the command line readers that implement the interactive actions of the suite. This should be the only package that interacts with the user terminal.

The vdmj.messages package contains values classes and exceptions for holding error and warning messages.

The vdmj.debug package contains classes which implement the DBGp protocol described in [6]. This is used by the Eclipse debugger in Overture to debug specifications using VDMJ.

The vdmj.utils package contains common utilities used by all other packages, as well as the native code used by the "stdlib" VDM libraries.

2. Package Detail

2.1. vdmj

Class Summary	
VDMJ	The main class of the VDMJ parser/checker/interpreter.
VDMPP	The main class of the VDM++ parser/checker/interpreter.
VDMSL	The main class of the VDM-SL parser/checker/interpreter.
VDMOV	The main class of the Overture VDM++ parser/checker/interpreter.
ExitStatus	An exit status code.
Settings	A class holding flags for -pre, -post, -inv and -dtc

The vdmj package contains the “main” abstract VDMJ class for the suite, plus two subclasses to parse, type check and interpret a specification in the VDM-SL or VDM++ dialect, and one class to use the Overture parser for VDM++.

These classes just collect together a sequence of other classes to parse, type check and interpret the specification, depending on the command line arguments. The following (working) example illustrates the principles, using VDMJ packages to create a minimal interactive VDM-SL program:

```
public static void main(String[] args) throws Exception
{
	File file = new File(args[0]);
	LexTokenReader ltr = new LexTokenReader(file, Dialect.VDM_SL);
	ModuleReader mr = new ModuleReader(ltr);
	ModuleList modules = mr.readModules();

	if (mr.getErrorCount() == 0)
	{
		TypeChecker tc = new ModuleTypeChecker(modules);
		tc.typeCheck();

		if (TypeChecker.getErrorCount() == 0)
		{
			ModuleInterpreter interpreter =
				new ModuleInterpreter(modules);
			interpreter.init(null);
			CommandReader reader =
				new ModuleCommandReader(interpreter, "$ ");

			List<File> files = new Vector<File>();
			files.add(file);
			reader.run(files);
		}
	}
}
```

The example would be almost exactly the same for VDM++ with "Class" substituted for "Module" in the various names (ClassReader, readClasses, ClassTypeChecker etc.).

VDMJ has a *-o* option which causes the parsed and type-checked specification to be written out to the filename specified by the argument to *-o*. Note that the entire tree is written to the file (if there are no type check errors), so if VDMJ is invoked with four specification files, and one *-o* option, the classes specified by all four files are written to the one output file. Any input file called "*.lib" (rather than the more normal *.vpp or *.vdm) is assumed to be a pre-compiled library created by *-o* and is loaded as

such without repeating the type checking. This can be faster for very large specifications. The example below shows the creation and use of IO.lib.

```
$ vdmpp -o IO.lib stdlib/IO.vpp
Parsed 1 class in 0.266 secs. No syntax errors
Type checked 1 class in 0.015 secs. No type errors
Saved 1 class to IO.lib in 0.125 secs.
```

```
$ vdmpp -i hello.vpp IO.lib
Loaded 1 class from IO.lib
Parsed 1 class in 0.422 secs. No syntax errors
Type checked 1 class in 0.015 secs. No type errors and 1 warning
Initialized 2 classes in 0.0 secs.
Interpreter started
> p new A().op()
Hello world!
= true
Executed in 0.015 secs.
```

2.2. Comments

The structure of these classes isn't very flexible. In particular, if the "load" command is given to the CommandReader, it is awkward to recover if parse/type errors are discovered in the new set of filenames (the List<File> passed to the run method is updated by the command reader to pass the new file names back to be parsed and checked).

The *-o* option was not as successful as I'd hoped. The idea was to serialize the tree after type checking, and load that back in quickly rather than re-parsing and re-checking the specification. Unfortunately, even though the serialization is passed through a gzip compression, the files produced are quite large and take a significant amount of time to decompress and de-serialize. For very small specifications it is faster to re-parse and check them, though for larger specifications there is an advantage.

Note that the serialization of the tree has to be for complete specifications (no unresolved references). This is because there is no link-editing available to combine partial specifications, and VDM++ does not have a way to identify externals (VDM-SL has module imports, but VDM++ has nothing similar). You can load multiple library files, or a mixture of library and source files.

2.3. vdmj.lex

Class Summary	
LatexStreamReader	A class to filter out LaTeX markup from an input file.
BacktrackInputReader	A class to allow checkpoints and backtracking while parsing a file.
LexBooleanToken	A class to represent a boolean token.
LexCharacterToken	A class to represent a character literal token.
LexIdentifierToken	A class to represent an identifier.
LexIntegerToken	A class to represent an integer literal token.
LexKeywordToken	A class to represent keyword tokens.
LexLocation	A class to hold the location of a token.
LexNameList	A class to hold a list of LexNameTokens.
LexNameToken	A class to hold a name.
LexQuoteToken	A class to represent a quote type token.

LexRealToken	A class to represent a real literal token.
LexStringToken	A class to represent a string literal token.
LexToken	The abstract parent class for all lexical token types.
LexTokenReader	The main lexical analyser class.

Enum Summary	
Dialect	An enumeration to indicate the VDM dialect being parsed.
Token	An enumeration for the basic token types.

Exception Summary	
LexException	An exception class for lexical analyser exceptions.

The vdmj.lex package contains all classes concerned with the lexical analysis of specifications. This includes the lexical token reader, plus a set of value classes for representing all types of lexical token.

The base class of the lexical system is BacktrackInputReader, which allows a stack of markers to be pushed within a stream of characters, returning the read pointer to the previous marker when a pop operation is performed. The class allows truly random movement within the stream – which is held as an array of Unicode chars, once it has been loaded. It opens input files with a LatexStreamReader object (extends InputStreamReader), which strips out LaTeX markup, while preserving the line numbers (turning LaTeX markup lines into blank lines).

LexTokenStream extends BacktrackInputReader. Its purpose is to make the input file look like a continuous stream of LexTokens. The nextToken returns the next token from the stream, and getLast will (repeatedly) return the last token read. Push and pop mark the stream and return to a mark respectively; unpush removes a marker without returning to it; the retry method does a pop followed by a push.

The LexLocation class is used throughout the system to represent a location within source code, for error message reporting. The toString method of the class produces a string which can be appended to another message:

```
"in <class/module> (<filename>) at <line>:<column>"
```

All value objects in the system which have a sensible position in the source code have a LexLocation associated with them. The class is also used to implement execution coverage tracking.

There are several subclasses of LexToken which represent tokens that contain a value which is more naturally represented by a Java primitive type. For example, LexRealToken includes a double field, and LexBooleanToken contains a boolean.

The Token enumeration is the basic label for all token types. The enumeration includes a lookup method which, together with a Dialect, decides whether a given string is a token or not. Note that the dialect affects this: “static” is a token in VDM++ but is a legal variable name in VDM-SL, for example.

Lexical analysis throws a LexException if there are any problems. Recovery is left to the syntax analysis layer.

2.3.1. Comments

To report accurate line positions, the token reader has to know the width of tabs. This is currently fixed at 4 characters by the TABSTOP field of LexTokenReader. It should probably be a settable field.

There is some modest ugliness concerned with the parsing of certain symbol sequences that look like other tokens, eg. “mk_mod`name”. Naively, this would be a name (a LexNameToken) with a module part of “mk_mod” and the identifier part of “name”, but actually this is parsed as a single identifier, so

that the syntax analyser can remove the “mk_” part and reveal the actual name.

Note that LexLocations have both start and end location information, though this is not used. The intention was to be able to identify blocks of source code that could be highlighted (eg. a whole statement or function, rather than just the start of one).

Recovering from a lexical error by raising an exception up to the syntax analysis may not be a good idea. The only recovery the syntax layer can do is to read up to some sort of safe point (eg. a semi-colon), and proceed from there. This gives comparatively poor error messages in general. It might be better to inject a likely token into the lexical stream, rather than throw an error and interrupt the stream.

LexNameTokens have a module/class name part and a simple name part (ie. they represent a grammatical *name* such as C`xyz). Whenever a name is created (as opposed to an identifier), it therefore has to include its class or module name, even if none was actually used in the specification (eg. simple parameter names are identifier patterns that are characterized by a name token, so all parameter names are held as LexNameTokens like C`x). That would be fine, except that in VDM++ the presence or absence of the class qualifier in a name can have semantic significance – explicitly identifying a member in a class hierarchy for example, with "object.X`name()". So LexNameTokens have an *explicit* flag, which means that the class name was explicitly specified by the caller, not implicitly filled in by the parser, and that flag is used during VDM++ type checking and runtime to identify the correct definition for the name. This works, but it may be over complicated. There are places where the *explicit* flag is set for very obscure reasons, which is a maintenance hazard. It may be better to take the Overture AST approach and allow names that simply don't have a class/module definition, and then take account of this when looking up definitions.

To enable VDM++ function/operation overloading, LexNameTokens optionally include a TypeList, qualifier representing the parameter types of the function or operation. The qualifier, if present, is used in the equals method of the class, and when searching for a name in a function/operation apply, the name sought is qualified with the actual types of the arguments. The equals function uses the TypeComparator to make the test, so a function declared with an int parameter will match one sought with a nat1 argument, etc. This system for managing overloading is not without its problems. Firstly, the use of the TypeComparator makes the equals method quite heavy. Secondly, it means that names cannot naively be used in Java maps because the hashCode of a function declaration name, and a function apply name may not be the same (if the argument types are not identical to the parameter types). Thirdly, it causes trouble when functions are applied via function variables (ie. lambda values) – such values can either be qualified with their parameter types or not, but since a function variable can either be applied (where argument types can be deduced) or just passed on (where they cannot), one or other of the name lookups will fail. To compensate, VariableExpression (which resolves simple names) will perform an unqualified "plain" name lookup if a qualified one fails. But a qualified lookup may succeed inappropriately by picking up an outer definition from the environment, when an inner unqualified name exists. This is tricky to solve.

2.4. vdmj.syntax

Class Summary	
Reader	The parent class of all syntax readers.
ExpressionReader	A syntax analyser to parse expressions.
TypeReader	A syntax analyser to parse type expressions.
DefinitionReader	A syntax analyser to parse definitions.
ClassReader	A syntax analyser to parse class definitions.
ModuleReader	A syntax analyser to parse modules.
PatternReader	A syntax analyser to parse pattern definitions.
BindReader	A syntax analyser to parse set and type binds.
StatementReader	A syntax analyser to parse statements.

Exception Summary	
ParserException	A syntax analyser exception.

The vdmj.syntax package contains all the syntax analysis classes. This includes eight “readers”, which implement a backtracking recursive descent parser, based on a stream of lexical tokens. The readers are all sub-classes of an abstract Reader class.

Every Reader subclass is constructed by being passed a LexTokenReader object. The reader is then responsible for returning one or more syntactic elements that it is designed to parse. For example, an ExpressionReader can be attached to a lexical stream, and be used to read an expression, or a comma separated expression list. Readers typically have methods called “read<something>” to preform the actual parse (eg. the minimal example above calls readModules from a ModuleReader).

Note that one type of reader usually needs other types of reader to complete its job. So for example, when a DefinitionReader is parsing an explicit function definition, it will use the raw lexical stream to read the function name, it will use a TypeReader to read the function’s type signature, a PatternReader to read the parameter patterns, and an ExpressionReader to read the body of the definition and any following pre or post conditions. All readers cache instances of the other readers used, and methods like getTypeReader either return the previous instance or create a new one. Note that there is no positional state information held in a reader therefore – it must depend on the LexTokenReader it contains to (say) determine the last token read.

Several parts of the VDM grammar cannot be parsed unambiguously by reading lexical tokens in a strict sequence. For example, several parts of the grammar use a <pattern bind> symbol, which is defined to be either a pattern or a bind, but it is not possible to distinguish a pattern from a bind by looking at the next token in the stream – a type bind is a <pattern>:<type> for example, so it looks like a pattern to start with but turns out to be a bind. To overcome this, the parsers are able to backtrack. That is, the start of the <pattern bind> is marked (a push operation on the lexical stream), and an attempt is made to parse the “longest” possibility, which is a bind in this case; if that fails, the stream is popped back to the marker, and the other possibilities are tried. If all possibilities fail, there is clearly an error, though it is not clear which branch contains the error (eg. is it a pattern that is malformed or a bind that is malformed?). In this case, the parser reports the error from the branch which managed to consume the most tokens after the marker before failing. This is assumed to be the most helpful error, though it is not certain what the user intended.

The following backtrack code pattern is used frequently by the readers. Note how the ParserException is used to carry “depth” information about how far the parser progressed, and the two depths are compared at the end to see which exception to throw. The code calling this method may itself be backtracking, having pushed its own markers in the stream.

```
public PatternBind readPatternOrBind()
      throws ParserException, LexException
{
      ParserException bindError = null;

      try
      {
            reader.push();
            Bind b = readBind();
            reader.unpush();
            return new PatternBind(bind.location, b);
      }
      catch (ParserException e)
      {
            reader.pop();
            e.adjustDepth(reader.getTokensRead());
            bindError = e;
      }

      try
      {
```



```
                reader.push();
                Pattern p = getPatternReader().readPattern();
                reader.unpush();
                return new PatternBind(p.location, p);
            }
            catch (ParserException e)
            {
                reader.pop();
                e.adjustDepth(reader.getTokensRead());
                throw e.deeperThan(bindError) ? e : bindError;
            }
        }
```

The top level of the parsers contain the recovery code to attempt to move the parse beyond a given syntax error. This is done by each top level case (eg. parsing a whole function definition) defining two lists of tokens: those which should be read up to, and those that should be read up to and past in the event of an error. Then a common recovery method (“report” in the Reader class) reads tokens up to one of those specified before continuing with the parse. The objective is, say, for a Statement reader to read tokens up to the end of the broken statement before continuing. In general this is not foolproof, and often syntax errors produce a short cascade of unrelated errors later in the specification.

2.4.1. Comments

There are some ugly parts to the parsing. One is concerned with equals definitions, which are defined as "def" <pattern bind>=<expression> "in" <expression>, but if the <pattern bind> is actually a set bind of the form “e in set S”, this parses as "s in set (S = <expression>)". There is some nifty footwork in the code to get round this (see readEqualsDefinition in the DefinitionReader).

Another ugly case is concerned with object call statements, which grammatically look like object apply designators as they end in ...<name>(args). So the parser reads an apply designator, then looks inside it to see whether the object being applied is a field designator or an identifier. The former is an object member invocation, the latter is a simple operation call.

The Reader base class provides a set of methods for reading and optionally advancing by one token. The differences between them are subtle (eg. advance and return the next token, or return the current token and advance), and I suspect the code could be cleaned up by reducing the number of options.

Recursive descent parsing always has difficulty with accurate error reporting and recovery. The method chosen is not perfect.

2.5. vdmj.ast

Class Summary	
ASTConverter	The AST tree converter.

The vdmj.ast package contains one class which is used to translate an Overture AST parse tree into the equivalent structure for VDMJ. The class is constructed with a filename and an IOmlDocument obtained from the OvertureParser. A convertDocument method converts the document tree into a list of ClassDefinitions.

2.5.1. Comments

This is a very large monolithic class, which should probably be broken up.

The Overture parser is compiled using Java 1.6, but the interface does not use generics – for example, the *get* methods to extract a list of classes from a specification is declared as a raw Vector,

not a List<IOmlClass>. The converter assumes that the get methods will be converted to use generics one day, and assigns their return values to (likely) generic values, like List<IOml...>. To get this to compile, there is a class-wide annotation to suppress unchecked warnings. When the AST interface is upgraded, this can be removed and the compiler should point out any type inconsistencies between the assumed interface and the actual declarations (though since the code works, this is very likely to be right).

2.6. vdmj.types

Class Summary	
Type	The parent class of all static type checking types.
***Type	A *** type. There are 25 such classes.
BasicType	The parent of the basic types (numbers, booleans and characters).
NumericType	The parent of the numeric types (real, rat, int, nat, nat1)
InvariantType	A type which has an invariant function associated with it.
NamedType	A type with a name.
OptionalType	An optional type.
ParameterType	A type associated with a polymorphic parameter name.
UnionType	A union of types.
UnknownType	A type representing a parser error.
UnresolvedType	A type name identifier by the syntax analyser.
VoidType	A type indicating the absence of a type.
VoidReturnType	A type indicating that a return statement has returned “()”.
PatternListTypePair	A pattern list combined with a single type.
PatternTypePair	A pattern plus a type.
TypeList	A list of types.
TypeSet	A set of types.

The vdmj.types package contains value classes that represent the various static types that can be contained in a VDM specification. They are all sub-classes of an abstract Type class.

Most simple types have a class dedicated to them of the same name, for example IntegerType or QuoteType. Similarly, the composite types have classes, like RecordType, SetType, SeqType and MapType; these have fields that in turn indicate the types of their components.

All types have a method to “resolve” themselves. The process of type resolution occurs early in the type checking process (see below), and turns UnresolvedTypes, which are just the names of types from the syntax analysis, into the actual type of the corresponding definition. The core of this happens in the typeResolve method of UnresolvedType, though other types call typeResolve recursively for any types that they contain – for example, FunctionTypes must resolve their parameter types and return type. The type resolution mechanism contains a recursive defence to avoid types which reference themselves from blowing the stack.

All types have a method to “polymorph” themselves. This means that given an actual type parameter, they substitute that type for any ParameterTypes they contain to yield a new Type object. This is used during the type check and execution of polymorphic function instantiations, when the actual type parameters are known.

All types implement a number of “is” and “get” methods – for example, (boolean) isMap and (MapType) getMap. These are used during type checking to determine whether a type is suitable for use in (say) a map application context. For simple types, these methods return false for the “is” method, except for the type concerned, which returns “true”; and “this” from the “get” method. For

more complex types, these methods support the situation where the static type checking cannot know the actual runtime type, but knows that it is one of several. In this case – the most obvious example being a UnionType – the “is” method will return true if any of the member types of the union would return true; and the “get” method will construct a new type representing the aspects of the applicable members of the union, all spliced together. The type checking can then proceed using the information of this single blend of possibilities, in the knowledge that the tests it is making could occur at runtime.

For example, if a type is a union of two records, each of which has a field called “label”, one of which is a “seq of char” and the other of which is a “nat1”, the getRecord method of the UnionType would return a new synthetic RecordType with a single field called “label”, with type “(seq of char) | nat1”. So the type checking of access to the label field would proceed as though that was its type, even though at runtime the type will be one or the other.

All types have a getAllValues method which is used during the evaluation of type binds. The method returns all the values for the type, though the only type which implements this is BooleanType; other types throw an exception if this is called.

UnknownTypes are used during error handling. Typically, an error will be encountered and reported, but rather than returning (say) the type of a sub-expression from type checking, an UnknownType is returned. This type has the property that all of its “is” methods return true, and its “get” methods will try to return a plausible Type. This means that subsequent type checking will not produce a cascade of errors as a single type checking fault deep in a specification winds its way out to the top level.

The VoidType is usually used to mean the absence of a type, so for example an operation which returns “()” would be represented by a VoidType. During the type checking of a sequence of statements in an operation, any statements which follow a “return” statement on an execution branch will be unreachable (a warning). So to distinguish this deliberate return of nothing from most statements which yield nothing, the VoidReturnType is used.

The TypeList and TypeSet classes implement lists and sets of Types, respectively. These are used in processing when a collection of types are encountered, and they must be turned into a single product type (TypeList) or a single union type (TypeSet). The ProductType and UnionType classes contain one TypeList and TypeSet, respectively.

2.6.1. Comments

There is no ReferenceType (compare with a ReferenceValue), yet several of the types do contain directly referenced types (like OptionalType and BracketType). It might be possible to simplify the hierarchy by adding one.

Do we call them products or tuples? I went for ProductTypes and TupleValues in the end.

2.7. vdmj.expressions

Class Summary	
Expression	The parent class of all VDM expressions.
***Expression	An expression of type ***. There are >100 of these.
BinaryExpression	The parent of all binary expressions.
NumericBinaryExpression	The parent of all numeric binary expressions (+, -, *, /)
BooleanBinaryExpression	The parent of all boolean binary expressions (and, or, <=>, =>)
UnaryExpression	The parent of all unary expressions.
ExpressionList	A list of Expressions.

The vdmj.expressions package contains value classes that represent the various types of expression that can be defined in a specification. They are all sub-classes of an abstract Expression class, which defines methods for an expression to be type checked and evaluated.

The typeCheck method is implemented by all expressions, and is passed an environment defining the variables and types in scope, together with a NameScope which identifies what sorts of names are accessible (eg. whether state values are in scope – they are for expressions in operations, but not in functions).

The typeCheck method returns a Type which indicates the result of evaluating the expression in the environment passed. So literal expressions simply return an appropriate type, like IntegerType. More complex expressions have to consider whether the definitions they contain affect the environment, whether those definitions have to be type checked, and whether the type check of any sub-expressions returns the expected result for the overall expression.

For example, consider a “forall” expression. This will contain a bind list of variables, and a predicate to evaluate for each (at runtime). The typeCheck method of ForAllExpression is as follows:

```
@Override
public Type typeCheck(
      Environment base, TypeList qualifiers, NameScope scope)
{
      Definition def = new MultiBindListDefinition(location, bindList);
      def.typeCheck(base, scope);
      Environment local = new FlatCheckedEnvironment(def, base);

      if (!predicate.typeCheck(
            local, null, scope).isType(BooleanType.class))
      {
            predicate.report("Predicate is not boolean");
      }

      local.unusedCheck();
      return new BooleanType(location);
}
```

A MultiBindListDefinition is a type of Definition (see below) which, when given the bind list for the forall expression, can expand the Environment passed in to make the names and types of the bind variables visible. Once created, the new definition is type checked to make sure the bindings contain no errors (for example, if the bind list contains a set bind, the expression representing the set must be a SetType).

A new FlatCheckedEnvironment is created to chain the new definitions onto the base Environment passed in, and this is used to type check the predicate of the forall expression. The return value of this is the Type of the predicate, which must be a boolean expression. The local environment is then checked to see whether all the names added to it by the bind list were actually used when type checking the predicate; any unused variables generate a warning. Lastly, this method returns a boolean Type, since a forall expression returns a boolean.

The typeCheck method on Expression is also passed a TypeList. This is used when trying to resolve name overloading during function and operation application. The typeCheck method of ApplyExpression starts by generating a TypeList from the typeCheck of each of the argument expressions it has. That may generate (say) [int, int, bool]. That list is then passed to the typeCheck method for the root of the apply (the thing being applied). If this root is a VariableExpression or a FieldExpression, the name of the variable or field is qualified with the list of types passed in, and this is used to find an overloaded name of a function or operation definition that has parameters whose types are compatible with the arguments (note, compatible with, not identical to). If it turns out that the variable or field is actually a map (which has no qualifiers), the search is repeated for the name without type qualification.

The other important method on Expressions is the eval method. This is called to evaluate the expression given the runtime Context (the runtime equivalent of an Environment). The method returns a Value object, which can represent any value in VDM. The eval method for PlusExpression is as follows:

```
@Override
public Value eval(Context ctxt)
```



```
{
	breakpoint.check(location, ctxt);

	try
	{
		double lv = left.eval(ctxt).realValue(ctxt);
		double rv = right.eval(ctxt).realValue(ctxt);

		return NumericValue.valueOf(lv + rv);
	}
	catch (ValueException e)
	{
		return abort(e);
	}
}
```

All Expressions and Statements contain a Breakpoint object, and the eval method of all expressions and statements call their breakpoint's check method at the start. This usually does nothing, unless a breakpoint is set, in which case execution stops and the DebuggerReader is invoked.

The PlusExpression evaluates its left and right hand sides, and converts the resulting Value objects to raw Java doubles. The result is a new Value, creating using the valueOf method of NumericValue, which will create the simplest type of NumericValue capable of holding the result of the addition – for example, this might be an IntegerValue or a NaturalOneValue or a RealValue.

Note that the code may throw a ValueException, and that this is caught within the eval method rather than being propagated. This can only occur in the conversion of the sub-expression results to doubles, or the construction of the resulting Value. ValueExceptions indicate problems while evaluating an expression, but they are caught and propagated using the abort method, which creates and throws a ContextException (a Java RuntimeException). This is done to distinguish between expected value errors – for example, trying to convert a Value to one of several types in a union, and failing before the right one is found – and serious errors, which should cause the system to halt.

The most complex evaluations are for functions and operations, but these are delegated to the corresponding FunctionValue and OperationValue classes. This is so that a function can be separately created (for example via a lambda expression) and applied. Operations cannot be created like this, but having operation values too means that, in VDM++, an object becomes a map of names to values – whether those are instance variable values or member operations and functions. The eval method of the ApplyExpression to call a function is therefore just:

```
	try
	{
		Value object = root.eval(ctxt).deref();

		if (object instanceof FunctionValue)
		{
			ValueList argvals = new ValueList();

			for (Expression arg: args)
			{
				argvals.add(arg.eval(ctxt));
			}

			FunctionValue fv = object.functionValue(ctxt);
			return fv.eval(argvals, ctxt);
		}
		else if (object instanceof OperationValue)
		...
```

All Expressions implement a method called findExpression. This has a line number parameter, and all implementations are responsible for returning themselves if they start on the line number, or recursing into their sub-expressions if they have them. This is used to set breakpoints on specific lines of a specification.

2.7.1. Comments

The implementation of name overloading may be problematic. A LexNameToken is optionally qualified with a TypeList, and the equals method uses the TypeComparator (part of vdmj.typechecker) to make a compatible comparison of the two names' type lists. Care must be taken when comparing names, especially during the "bootstrap" phases when qualifiers are not yet available.

2.8. vdmj.statements

Class Summary	
Statement	The parent class of all statements.
***Statement	A statement of type ***. There are 40 or so of these.

The vdmj.statements package contains value classes that represent the various types of statement that can be defined in a specification. They are all sub-classes of an abstract Statement class, which defines methods for a statement to be type checked and executed.

Many of the principles for Statements are the same as those for Expressions covered above. Like Expressions, all Statements include a typeCheck method which is passed an Environment and a NameScope – though note that there is no need for a TypeList of qualifiers because an operation call can only be rooted on something that is already known by the statement (an operation name or an object designator), whereas function application in an expression has to evaluate an arbitrary expression to generate the root to which to apply the arguments.

Similarly, like Expressions, all Statements define an eval method which executes them, returning a Value – though statement executions usually return VoidValues.

Statements implement a method called exitCheck, which explores the statement tree (following blocks and branches from compound statements) looking for statements which can raise an exit status, like the ExitStatement itself. This is used in the type checking of statements that catch and process exits, like TrapStatements. The method returns a set of exit Types that can be thrown.

All Statements implement a method called findStatement. This has a line number parameter, and all implementations are responsible for returning themselves if they start on the line number, or recursing into their sub-statements if they have them. This is used to set breakpoints on specific lines of a specification.

2.8.1. Comments

The exitCheck is not perfect. In particular, the check does not cross the boundary of an operation call from a statement block, nor can it follow an expression evaluation that involves operation calls. That would require code to work out which operation(s) are actually involved, and it would require recursive defence against operations which recurse. So exitCheck produces false negatives (indicates that operations can't exit, when in fact they can). I gather VDMTools is better, but not perfect either. It is known to produce false positives.

2.9. vdmj.patterns

Class Summary	
Pattern	The parent type of all patterns.
****Pattern	A pattern of type ****. There are 14 such pattern types.
Bind	The parent class of SetBind and TypeBind.
MultipleBind	The parent class of MultipleSetBind and MultipleTypeBind.

PatternBind	A pattern or a bind.
PatternList	A list of patterns.

The vdmj.patterns package includes value classes to represent the various patterns and binds in a specification. They are all subclasses of an abstract Pattern, Bind or MultipleBind class.

Patterns generate definitions, given a Type. For example, the pattern “[a,b,c]” will produce definitions for the three integer variables, given that it is of type “seq of int”. The process of definition generation is recursive over the tree of nested patterns that might be defined. So the getDefinitions(Type) method which all patterns implement, will recurse for those pattern types which are defined as containing sub-patterns. The leaves of the pattern tree are the simple pattern types: BooleanPattern, CharacterPattern, etc. At the leaves, it is only IdentifierPattern which produces variable definitions.

Similarly, patterns generate a list of name/value pairs given a Value. The value is matched against the “shape” of the pattern and its sub-patterns, and any IdentifierPatterns at the leaves are populated with the corresponding part of the original value. This getNamedValues method is called from definitions which include patterns (like “values” definitions) when the actual values of the definition are required.

Patterns can also be asked for a simple list of their variable names, which involves a depth search for IdentifierPatterns.

Patterns include a typeResolve method because ExpressionPatterns can contain referenced to UnresolvedTypes that need to be resolved early in the type check.

A Bind, which is sub-classed by SetBind and TypeBind, comprises a pattern and a set of Values (either an explicit set, or the set of all the Values that a Type can generate, in theory). These are used in quantified expressions, like “exists {a,b} in set S & a.type = b.type”. Here the set pattern {a,b} contains two identifier patterns that must (potentially) iterate through all the elements of S, taking the corresponding values. To drive the iteration, the Bind needs to generate all the possible Values, which are then given to the pattern to generate name/value pairs for each iteration. Therefore Bind has a method called getAllValues, which is implemented by both sub-classes (the TypeBind implementation calls the getAllValues method of the Type, which is an error for everything except BooleanType).

MultipleBind, which is sub-classed by MultipleSetBind and MultipleTypeBind, is very similar except that they comprise a list of patterns and a set or type. But they still have a getAllValues method which collects together all the possible values in the set.

Note that in VDMJ, the generation of all values from a set includes the permutations of all the orderings of the values in that set. Internally, sets of Values are held in a ValueSet which is actually an ordered list (but with set semantics, with regard to no duplicates). Therefore the getAllValues methods for the various set binds call the permuteSets method of ValueSet (via the same method on SetValue). Note also that the original set is sorted before this process, which means that given the same set content, the order of processing in a bind (and therefore any looseness based on it) is consistent.

```
@Override
public ValueList getBindValues(Context ctxt)
{
	try
	{
		ValueList results = new ValueList();
		ValueSet elements = set.eval(ctxt).setValue(ctxt).sorted();

		for (Value e: elements)
		{
			e = e.deref();

			if (e instanceof SetValue)
			{
				SetValue sv = (SetValue)e;
				results.addAll(sv.permutedSets());
			}
			else
			{
```



```
                        results.add(e);
                    }
                }

                return results;
            }
            catch (ValueException ex)
            {
                abort(ex.getMessage(), ctxt);
                return null;
            }
        }
```

2.9.1. Comments

Binds and MultipleBinds look like they have a lot in common and could probably be put together to avoid a small amount of code duplication.

2.10. vdmj.traces

Class Summary	
TraceDefinition	An abstract class representing a trace definition.
TraceDefinitionTerm	A class representing a sequence of trace definitions.
TraceLetBeStBinding	A class representing a let-be-st trace binding.
TraceLetDefBinding	A class representing a let-definition trace binding.
TraceRepeatDefinition	A class representing a repeated trace definition.
TraceCoreDefinition	Abstract of all core trace expressions.
TraceApplyExpression	A class representing a core trace apply expression.
TraceBracketedExpression	A class representing a core trace bracketed expression.
TraceNode	An abstract class representing an expansion node.
AlternativeTraceNode	An expansion node for alternatives.
RepeatTraceNode	An expansion node for repeats.
SequenceTraceNode	An expansion node for sequences.
StatementTraceNode	An expansion node (leaf) for statement applies.
TestSequence	A sequence of CallSequences
CallSequence	A sequence of CallObjectStatements.
Permutor	A utility to permute a set of values.

Enum Summary	
Verdict	A test outcome: PASSED, FAILED or INDETERMINATE.

The vdm.traces package includes classes to represent the definitions which can occur in a VDM++ "traces" section, and their subsequent expansion and execution.

The subclasses of TraceDefinition are uncontroversial and follow the structure of the trace grammar closely. These classes have the usual typeCheck methods which permit the trace specifications to be checked as part of the overall specification type check phase.

In order to evaluate traces, they must first be expanded into all the possible execution paths

represented by the definition. For example, a trace that is of the form "a;(b|c);d" would expand to "a;b;d" and "a;c;d". Similarly, "a{1,3}" would expand to "a", "a;a" and "a;a;a". Repeats, sequences and alternations multiply together to generate large numbers of tests very quickly – this is the whole point of combinatorial test specification. The TraceNode class and its subclasses represent the expanded traces, and are generated by "expand" methods on all TraceDefinitions. Strictly, these classes don't expand the traces, but produce a tree structure that is capable of expanding them. The getTests method of TreeNodes actually expands the tests, returning a TestSequence, which is a list of CallSequence (ie. a list of tests), and a CallSequence is a list of operation applies (of CallObjectStatements).

A NamedTraceDefinition (see below) is created for each parsed trace definition, and this produces an operation in the class which, when executed, will expand the trace and execute all of the tests in turn. Between each test execution the system is initialized (the init method of the ClassInterpreter) and a new object is created to run the test. The body of these operations are TraceStatements, which are constructed with reference to their NamedTraceDefinition. The eval method of this class first gets a TestSequence from the NamedTraceDefinition. Then it creates an Environment which is suitable to type check the statements in the tests – it is necessary to type check statements before their execution. Then for each CallSequence in the TestSequence, it initializes the ClassInterpreter, and calls its runtrace method, passing the CallSequence and the Environment. The runtrace method executes the test and returns a list of java.lang.Object, being either the return values from the test steps or error messages, and the last item will always be an instance of a Verdict object, indicating the test outcome. Tests which have a FAILED verdict are "stemmed" and then remaining tests in the TestSequence which have the same stem are marked as "filtered" by this test – ie. there is no point in running them because their initial sequence of calls will fail at the same point, for the same reason. Before each tests is executed, its filtered flag is tested, and such tests are not executed.

2.10.1. Comments

There is a minor quibble in the parsing of trace sections, in that the grammar prohibits the use of semi-colons between trace definitions (while permitting them to separate parts of a trace definition). VDMJ permits these separate semi-colons; the Overture parser currently does not.

I tried very hard to combine the expansion of the tests with the classes that represent the parsed trace definitions – this is just a tree after all. But I couldn't get it working properly, hence the solution where the definitions are expanded into a separate tree, which is then "walked" to generate the trace permutations.

Filtering is done by creating the toString of the failed test stem, and String.startsWith of the toStrings of the remaining tests. This creates the toStrings many times and isn't very efficient.

2.11. vdmj.definitions

Class Summary	
Definition	The abstract parent of all definitions.
AccessSpecifier	A class to represent a [static] public/private/protected specifier.
AssignmentDefinition	A class to represent assignable variable definitions.
ClassDefinition	A class to represent a VDM++ class definition.
ClassInvariantDefinition	A VDM++ class invariant definition.
EqualsDefinition	A class to hold an equals definition.
ExplicitFunctionDefinition	A class to hold an explicit function definition.
ExplicitOperationDefinition	A class to hold an explicit operation definition.
ExternalDefinition	A class to hold an external state definition.
ImplicitFunctionDefinition	A class to hold an implicit function definition.

ImplicitOperationDefinition	A class to hold an explicit operation definition.
ImportedDefinition	A class to hold an imported definition.
RenamedDefinition	A class to hold a renamed import definition.
InheritedDefinition	A class to hold an inherited definition in VDM++.
InstanceVariableDefinition	A class to represent instance variable definitions.
LocalDefinition	A class to hold a local variable definition.
MultiBindListDefinition	A class to hold a multiple bind list definition.
MutexSyncDefinition	A class to hold a mutex synchronization definition.
PerSyncDefinition	A class to hold a permission synchronization definition.
StateDefinition	A class to hold a module's state definition.
ThreadDefinition	A class to hold a thread definition.
TypeDefinition	A class to hold a type definition.
UntypedDefinition	A class to hold a definition of, as yet, an unknown type.
ValueDefinition	A class to hold a value definition.
NamedTraceDefinition	A class to hold a named trace definition.
ClassList	A class for holding a list of ClassDefinitions.
DefinitionList	A class to hold a list of Definitions.
DefinitionSet	A class to hold a set of Definitions with unique names.

The vdmj.definitions package contains classes representing the definitions in a specification. They are all sub-classes of an abstract Definition class.

All definitions have a few things in common (fields of the abstract class). They belong to a Pass, which guides the type checking; they have a location; they have a name, though this may be null if they contain sub-definitions; and they have a NameScope to define what sort of name(s) they define, which compliments the name scope used in type checking that searches for names of certain types.

Definitions define a typeResolve method which is used very early on to resolve the UnresolvedTypes that may have come through from the syntax analysis. For example, an ExplicitFunctionDefinition would typeResolve the Type of the function, and if there were pre or postconditions, these expressions would be typeResolved, and any parameter patterns would be typeResolved.

Definitions also define a typeCheck method, which is similar to the ones defined for Expression and Statement, except that there is no Type to return. For example, the ExplicitFunctionDefinition performs the following tasks in its typeCheck method:

- If there are any polymorphic type parameters for this function, check that the overall function type does not reference any type parameters except those named type parameters.
- For each type parameter, create a LocalDefinition of a ParameterType and add this to a local Environment.
- Check that the parameter patterns match the overall Type’s parameters, and iterate through curried sets of parameters, using the return value from the overall Type (and its return value and so on for subsequent sets of parameters). Remember the expected result.
- Extend the local Environment with definitions for all the variables of all the patterns from all of the curried parameter sets.
- Type check the definitions this produced in the base environment (this will just do type resolution, if necessary).

- Label the local Environment as static (VDM++) if the definition's access specifier is static.
- If we are in VDM++ and the function is not static, add a "self" definition to the local Environment.
- If there is a precondition expression, type check the definition for it.
- If there is a post condition expression, type check the definition for that too.
- Type check the body expression of the function, remembering the actual type returned.
- If the actual return type is not assignable to the expected return type, raise an error.
- If the VDM++ accessibility of the expected return type is narrower than that of the definition itself, raise an error (eg. a public function cannot have a private return type).
- If the function is recursive and does not define a "measure" function, raise a warning, else if there is a measure defined, check that it exists and has the correct type.
- Check that the parameter variables have been referenced in the local Environment, else raise an unused parameter warning. (This is suppressed for pre and post conditions, which are permitted to not necessarily use their implicit parameters).
- Return.

This illustrates the principles that are used by all definitions' typeCheck methods.

Some definition types are only used to "wrap" others. For example, during module imports and exports, a definition may be imported and/or renamed. These methods just delegate their calls to the referenced definition that they wrap.

As with Patterns, definitions can yield their contained definitions or a list of names or name/value pairs that they define. This is the purpose of the getDefinitions, getVariableNames and getNameValuePairs methods. Simple local definitions only define one variable, but many definition types include a Pattern specifier that may define many variables.

The findName method is implemented by all definitions to return whether they define a name being sought by type checking. As above, for simple definitions, this just compares their name and scope with that being searched for, but for definitions that include patterns, all the names generated by the pattern must be considered.

The findType method is implemented by those definitions that define a type (TypeDefinition, StateDefinition and ClassDefinition).

Definitions also define findExpression and findStatement methods which recurse into their bodies in search of an expression or statement that starts on the given line.

2.11.1. Comments

There is a great deal of commonality between some definitions, especially implicit and explicit functions and operations. It might be possible to simplify the code by creating abstract bases for these.

2.12. vdmj.modules

Class Summary	
Export	The parent class of all export declarations.
Export***	A class for representing exports of a given type.

Import	The parent class of all import declarations.
Import***	A class for representing imports of a given type.
Module	A class holding all the details for one module.
ModuleList	A list of Modules.

The vdmj.modules package contains value classes that represent the modular structure of VDM-SL specifications, including their import and export declarations.

The only purpose of these classes is to represent the parsed module structures from the specification, and to generate/find the list of exported and imported definitions that extend the scope of what is visible from a single module.

The ModuleList class is important because it contains the "initialize" method which is used to set the initial state of all modules when the interpreter is started.

2.13. vdmj.typechecker

Class Summary	
TypeChecker	The abstract root of all type checker classes.
ClassTypeChecker	A class to coordinate all class type checking processing.
ModuleTypeChecker	A class to coordinate all module type checking processing.
Environment	The parent class of all type checking environments.
FlatEnvironment	Define the type checking environment for a list of local definitions.
FlatCheckedEnvironment	Define the type checking environment for a list of local definitions, including a check for duplicates and name hiding.
ModuleEnvironment	Define the type checking environment for a modular specification.
PrivateClassEnvironment	Define the type checking environment for a class as observed from inside.
PublicClassEnvironment	Define the type checking environment for a set of classes, as observed from the outside.
TypeComparator	A class for static type checking comparisons.

Enum Summary	
NameScope	An enum to represent name scoping.
Pass	An enum to indicate which type checking pass a definition belongs to.

The vdmj.typechecker package includes classes which organize the type checking of VDM-SL and VDM++ specifications. Most of the actual type checking is performed by the definition, expression and statement classes that collectively describe the specification (above), but the typechecker package defines the supporting classes to invoke the type checking methods of the other objects, and to represent the static environment in which type checking is performed.

Type checking is different for VDM-SL and VDM++, though the checking of the basic statements and expressions is very similar. Therefore there is a common abstract TypeChecker class with two subclasses to deal with VDM_SL and VDM++: ModuleTypeChecker and ClassTypeChecker. The subclasses are constructed with a list of modules or classes – which is the overall result of a successful syntax analysis – and they implement a single abstract method from their parent, called typeCheck. The method takes no arguments and returns no result. Any errors or warnings raised during type checking are recorded by the parent class (see VDMMessage).

The sequence of events is slightly different for the type check of modules and classes, but they follow the same principles. For modules, the sequence is:

- Check for duplicate module names in the list passed
- For each module, generate its definitions' implicit definitions (like pre and post functions)
- For each module, check the export definitions exist and are of the declared type, and make a list of exported definitions for the module.
- For each module, go through the import definitions and resolve against the exports.
- Create a list of all definitions from all modules (including their imports), create an Environment that contains them all, and attempt to perform type resolution on them – ie. find the type definition for every named type.
- In the pass order: [types, values, definitions], for each module, create a ModuleEnvironment representing the visible definitions, and type check the definitions of the given pass. This calls the typeCheck method on the definitions, which calls the similar method on the definitions subparts, if any.
- Report any discrepancies between the final checked types of the modules' definitions and their explicit imported types.
- Any definition names that have not been referenced or exported produce "unused" warnings.

There are a couple of important points to note:

Firstly, the syntax analysis does not understand anything about the relationship of a type name in a declaration to its type definition. All type names come through from the syntax phase as UnresolvedTypes, which simply have a name. So a very early phase of the type checking must find the corresponding type definition, in order to understand the structure of the type and what is/is not a legal type manipulation. This is done on a global basis, even though not all types are in scope for a module (all type names are fully qualified with a module name, so there is no ambiguity). A subsequent pass, which uses just those definitions that are visible, will subsequently spot any scope problems.

Secondly, "environments" are used to support the type checking. An environment (a subclass of the abstract Environment class) is essentially a list of names and corresponding definitions that are in scope at any point. The different subclasses allow the different scope rules to be followed, so for example a module's type checking will start with a ModuleEnvironment that references a single module definition, and understands the rule about resolving names from its imported definitions and its own definitions. Different environments are then chained together, so when the module environment is passed to (say) a function definition, the type checking creates a new environment with local definitions for the names that are generated by the parameter patterns. Since the parameter names are in scope for the body of the function, the chain of two environments is passed to the type checking of the body expression. That may in turn involve "let" expressions that define further local variables that are chained onto the environment before their body is type checked, and so on. As the chain unwinds (as typeCheck methods return), each environment in the chain is checked to see whether all of its definitions were referenced; any that were not referenced generate "unused" warnings.

The two most important methods on an Environment subclass are findName and findType, which lookup definitions by name (using the same methods on the Definition classes they reference). There is also a name scope parameter passed to findName to indicate what sorts of names are in scope – for example, functions "see" value and parameter names, operations see these names and state variables, and the post conditions of operations see names, state and "old" names. The name scope (mask) is passed around the tree of typeCheck invocations as the names that are visible in a given context is generally only known by the caller – eg. an expression does not know that it is part of an operation, so it must be told from the outside that state variables are in scope.

Very similar principles are followed by the ClassTypeChecker, which performs the following actions:

- Make sure there are no duplicate class definitions.
- For all classes and their definitions, generate the implicit definitions. This includes the construction of the class type hierarchy and the implicit local names for access to inherited definitions.
- Create a public class environment that can see all public class definitions.
- For each class, chain a private class environment to the public environment, and perform type resolution on the definitions in the class.
- For each class, check for overloading and overriding of its definitions.
- In the pass order: [types, values, definitions], for each class, create a private class environment, and type check the definitions of the given pass.
- Check for any definition names that have not been referenced or exported, and produce "unused" warnings.

The use of public and private class environments to control the resolution of names is exactly analogous to the module case, though the class versions use the static/public/protected/private definition modifiers to decide on visibility.

The TypeComparator is used at the heart of type checking. The "compatible" method is used to decide whether two Types are assignment compatible (with "possible" semantics). This involves finding the “underlying” type (eg. removing the names of types) and making flexible recursive comparisons where unions are involved. It also has to have recursive defence, since type structures may reference themselves. Two optional types are always considered compatible (as they could both be nil). Compound types that involve more than one subtype (sets, sequences, maps, functions and operations, records and classes) are unpicked and their subtypes recursively compared. Finally, a Type.equals comparison is made between simple types.

The TypeComparator is also involved in proof obligation generation for subtype testing (see below). The "isSubType" method takes two types and returns a boolean indicating whether the first is a subtype of the second – for example, a nat1 is a subtype of a real (all nat1 values are real values), but not the other way round. With union types, a simple type is a subtype of a union if it is a subtype of any of the union members; a union type is a subtype of another union if every member of the first union is a subtype of the second union.

2.13.1. Comments

I had a lot of trouble getting the order of the initialisation and type checking right to be able to deal with all the specifications in the test suite. I’m not 100% sure that there aren’t still obscure orderings of declarations that will defeat it.

The TypeComparator is currently a static class, which means that its methods need to be synchronized to work with VDM++ threads (it has state for recursion defence). This isn’t really necessary, but there are quite a few places where the code would have to create a new TypeComparator instance if this is changed.

Type checking error messages are produced by static methods on the TypeChecker class, and the error count is held statically there too. This is because it is otherwise difficult to find the type checking object instance deep in a typeCheck call chain. This is in contrast to the syntax analysers, where a Reader keeps track of the errors for itself (plus any from the Readers it creates).

2.14. vdmj.pog

Class Summary

ProofObligation	The abstract root of all proof obligations.
***Obligation	A particular type of proof obligation.
ProofObligationList	A list of proof obligations.
POContext	The abstract root of all obligation contexts.
PO***Context	A particular type of obligation context.
POContextStack	A stack of obligation contexts.

Enum Summary	
POType	An enumeration of the various proof obligation types.

The vdmj.pog package defines classes that support the generation of proof obligations. Most of the actual obligation generation is performed by Definitions, Expressions and Statements, which include a getProofObligations method that returns a ProofObligationList.

A typical proof obligation contains a stack of nested "contexts" in which a particular obligation must be determined, plus a specific obligation to check. Therefore there are two distinct class hierarchies in the pog package: the POContext hierarchy, and the ProofObligation hierarchy.

For example, if "m" is a map of int to int, a simple function like:

```
f: int -> int
      f(i) == if i < 10 then m(i) + 1 else m(i) - 1;
```

would generate two proof obligations, requiring that the two m(i) map applications are correct in the "then" and "else" branches, respectively:

```
A`f(int): map apply obligation in 'A' (test.vpp) at line 15:32
(forall i:int &
  ((i < 10) =>
    i in set dom m))

A`f(int): map apply obligation in 'A' (test.vpp) at line 15:46
(forall i:int &
  (not (i < 10) =>
    i in set dom m))
```

Notice that both obligations contain an outermost "forall" that represents the possible function parameter values, and that the "if" test value is determined to be true or false in order to check the map application in the two branches. These two form the "context" of the proof obligation; the last line in both POs is the actual obligation, which tests that the argument is within the domain of the map.

The outer forall context is represented by a POFunctionDefinitionContext object, and the if/else contexts are represented by POImpliesContext and PONotImpliesContext objects, respectively. The proof obligation itself is a MapApplyObligation. The two contexts form a stack for each obligation, and these are held by a POContextStack, which extends Stack<POContext>.

To generate proof obligations for an entire specification, an empty POContextStack is created and the getProofObligations method of each definition is called, passing the stack. Depending on the definition type, the implementation may add to the context (eg. a function definition would push a POFunctionDefinitionContext) before calling getProofObligations for their inner expression(s) or statement(s). In the example above, the getProofObligation method of the IfExpression that comprises the function body would be called. That in turn would push an POImpliesContext before generating obligations in its "then" sub-expression, **popping** the stack, and pushing a PONotImpliesContext before generating obligations for everything in the else-if list, and final else. Lastly it would pop the stack back to the state it found it in before returning a list of all the generated POs.


```
public ProofObligationList getProofObligations(POContextStack ctxt)
{
	ProofObligationList obligations = ifExp.getProofObligations(ctxt);

	ctxt.push(new POImpliesContext(ifExp));
	obligations.addAll(thenExp.getProofObligations(ctxt));
	ctxt.pop();

	ctxt.push(new PONotImpliesContext(ifExp));	// not (ifExp) =>

	for (ElseIfExpression exp: elseList)
	{
		obligations.addAll(exp.getProofObligations(ctxt));
		ctxt.push(new PONotImpliesContext(exp.elseIfExp));
	}

	obligations.addAll(elseExp.getProofObligations(ctxt));

	for (int i=0; i<elseList.size(); i++)
	{
		ctxt.pop();
	}

	ctxt.pop();

	return obligations;
}
```

Notice that the IfExpression does not actually generate any proof obligations itself; it only sets up context so that obligations generated from its sub-expressions will be correct. The ApplyExpression actually generates the proof obligations in this example:

```
public ProofObligationList getProofObligations(POContextStack ctxt)
{
	ProofObligationList obligations = new ProofObligationList();

	if (type.isMap())
	{
		MapType m = type.getMap();
		obligations.add(
			new MapApplyObligation(root, args.get(0), ctxt));

		Type atype = argtypes.get(0);

		if (!TypeComparator.isSubType(atype, m.from))
		{
			obligations.add(new SubTypeObligation(
					args.get(0), m.from, atype, ctxt));
		}
	}
	...
```

Here, the apply expression is first tested for whether it is a map application (it could be a function or operation application), and if so, a new MapApplyObligation is created, which is passed the context. Similarly, if the apply argument type is not a subtype of the domain of the map, a SubTypeObligation is also generated.

The constructor of the ProofObligations generated use the context passed to generate the "value" of the obligation (ie. the string form of its expression):

```
public MapApplyObligation(
	Expression root, Expression arg, POContextStack ctxt)
{
```



```
        super(root.location, POType.MAP_APPLY, ctxt);
        value = ctxt.getObligation(arg + " in set dom " + root);
    }
```

The getObligation method on the context stack will generate a string composed of the contexts in the stack, plus the string passed in for the obligation. It is also responsible for the indentation and bracketing of the expressions. All ProofObligation subclasses are similar to the example above, though the more complex ones take a lot of effort to generate the string of the obligation. The most complex obligation is SubTypeObligation, which has a recursive private method to generate all the subtype tests that are required for the "structure" of the type being considered.

The process of proof obligation generation is exactly analogous with operation definitions which include statements – though often, statement obligations are generated without any context since in general it is too difficult to determine the scope of side effects generated by a specification.

2.14.1. Comments

The proof obligations generated are based on those in the CSK POG test suite. It is possible that there are other obligation types that should be added, especially in the case of VDM++.

ProofObligation subclasses generate the entire string of the obligation, including the context in which it is generated. This seemed better than keeping the context objects with the obligation, but it does mean that the structure of the PO is lost in the flat string.

The generation of expression strings depends on the accuracy of the toString methods for Expressions. These have been tidied up, but unfortunately preserving the precedence of the original operators means that many expression strings end up being excessively bracketed.

2.15. vdmj.runtime

Class Summary	
Interpreter	An abstract VDM interpreter.
ModuleInterpreter	The VDM-SL module interpreter.
ClassInterpreter	The VDM++ interpreter.
Context	A class to hold runtime name/value context information.
RootContext	An abstract context for the root of a function or operation call.
ObjectContext	A root context for object member invocations.
StateContext	A root context for non-object member invocations.
Breakpoint	The concrete root of the breakpoint hierarchy.
Stoppoint	A breakpoint where execution must stop.
Tracepoint	A breakpoint where something is displayed, but execution continues.
SourceFile	A class to hold a source file for source debug output.
ThreadState	A class to hold runtime information for a VDM thread.
VDMThread	A class representing a VDM thread.
VDMThreadSet	A class containing all active VDM threads.

Exception Summary	
ContextException	A fatal interpreter error, including a “stack” context to dump.
DebuggerException	An exception used to stop the interpreter cleanly from the debugger.

ExitException	An exception used to implement exit statements.
PatternMatchException	A non-fatal exception indicating a pattern match has failed.
StopException	An exception passed between threads to cause them to abort.
ValueException	A non-fatal exception concerning an evaluation.

The vdmj.runtime package defines the abstract interpreter, and its subclasses to interpret VDM-SL and VDM++ specifications. It includes classes to represent the runtime execution context, exceptions which can be generated at runtime, and the debugging and thread control classes.

At its simplest, an interpreter has to create a runtime environment that represents the globally visible name/values in a specification, then evaluate a function or operation indicated by the user, returning the result.

The runtime name/value pair environment is held by the Context class and its sub-classes. A Context extends a HashMap<LexNameToken, Value>, so it is capable of storing and retrieving named values. In the same way that Environment objects were chained together during type checking, Context objects can be chained together as evaluation creates new named values, and those names later disappear from scope.

Contexts allow named values to be retrieved by searching the present context, and then subsequent chained contexts. Hence a context for global variables might be chained with one for a function's parameters, and a further one defining variables in a "let" expression. Then a lookup of a variable would search this chain in reverse order.

In the case of functions or operations calling other functions or operations, the name searching should not proceed down the context chain beyond the nearest function or operation point – ie. if func1 calls func2, func1's variables are not in scope in func2. But global variables are visible in this case. Therefore a sub-class of Context, called RootContext is used to represent points in the context chain where the search should "jump" down to the global level. In fact there are two sub-classes of RootContext, one of which, ObjectContext, is specialized to old an object value for "self" (which is in scope at the start of object member calls), and the other, StateContext, which is able to have a module's state data attached (the actual state that is visible changes as the operation call stack jumps from module to module, so this must be held in the context chain somewhere).

To create the global environment, the interpreter asks the default module or the ClassList passed to create the name/values from their definitions. The ClassList initialize method will dump all class' public static values into one global public static Context object; while the ModuleList initialize method will let each module initialize itself. The ClassInterpreter will then take the global public static scope and add the private static scope of the default class before starting. The ModuleInterpreter takes the initial context from the default module before starting.

The code for the two interpreters' execute methods are similar therefore. Here is ClassInterpreter's:

```
@Override
public Value execute(String line) throws Exception
{
	LexTokenReader ltr = new LexTokenReader(line, Dialect.VDM_PP);
	ExpressionReader reader = new ExpressionReader(ltr, true);
	reader.setCurrentModule(getDefaultName());
	Expression expr = reader.readExpression();

	Environment env = new PublicClassEnvironment(classes);
	TypeChecker.clearErrors();
	expr.typeCheck(env, null, NameScope.NAMESANDSTATE);

	if (TypeChecker.getErrorCount() > 0)
	{
		throw new Exception("Type checking errors");
	}

	mainContext = new StateContext(
		defaultClass.name.location, "global static scope");
```



```
        mainContext.putAll(initialContext);
        mainContext.put(defaultClass.getPrivateStaticValues());
        mainContext.threadState.init();

        return expr.eval(mainContext);
}
```

Note that the method is given a string expression to evaluate, so first it parses and type checks the expression given. The last five statements contain the important bit: a new RootContext is created, the global public static content is added, the private statics of the default class are added, the threads system is initialized, and lastly the expression passed is evaluated in the context constructed. The return value is the result of the execute method.

Breakpoints are objects that exist in all expressions and statements – ie. inside anything that is directly executed. Breakpoint is the main class, with two sub-classes: Stoppoint and Tracepoint. All breakpoints have a “check” method which is called at the start of every expression and statement execution (and so must be fast!). The method is responsible for deciding whether to stop, and also recording that the statement or expression was reached for code coverage. In the case of a Stoppoint, the code must stop if it has no “test” expression, or the test evaluates to true (this is a user-inserted breakpoint); a Tracepoint will not stop, but it must evaluate its “show” expression and display it; and a regular Breakpoint will only stop if we are single-stepping. If the breakpoint decides to stop, it instantiates a DebuggerReader class, which is responsible for interacting with the user via the command line (alternatively, it interacts with the current DBGPReader object – see 2.18). As users set and clear breakpoints, the affected expression or statement is found (using findExpression) and the breakpoint member of the Expression or Statement object affected is changed by DebuggerReader. The interpreter manages the list of current breakpoints set, and the DebuggerReader has a reference to the interpreter. This also enables the DebuggerReader to evaluate “ad hoc” expressions when a stop point is reached. The SourceFile class enables selected source files to be read and lines displayed for clearer single stepping and breakpoint management, as well as displaying test coverage output.

The DebuggerReader’s run method is internally synchronized on the DebuggerReader class. This is to prevent more than one thread (in VDM++) from gaining control of the console. It unfortunately means that single stepping can have confusing results if many threads have actually stopped and are waiting for access to the console. See the User Guide for more information about this.

A StartStatement creates new VDMThread classes (which extend java.lang.Thread) to execute the VDM threads started. These create new ObjectContexts to form the root of the thread based on the object instance started, and otherwise perform a normal execution of the “thread” statements defined in the class, via a synthetic ThreadStatement. VDMThreads add themselves to the set managed by VDMThreadSet, and this also provides methods to suspend, resume or abort all threads in one go, which is used by the debugger.

Context objects include a ThreadState field, a reference to which is copied as Contexts are chained together, and the initial value of which is set when execution of a thread starts. The ThreadState tells the breakpoint system whether the current thread is suspended or single-stepping etc., and is used in the check method, and is manipulated by the DebuggerReader.

Thread scheduling is left entirely to the Java thread scheduler. All threads have the same priority.

2.15.1. Comments

I can’t remember why we can’t use initialContext directly to run things. It might just be that as the default class is changed, we want to dump different private statics into it.

2.16. vdmj.values

Interface Summary

ValueListener	Implemented by classes that watch for changes to a Value.

Class Summary	
Value	The parent of all runtime values.
****Value	A value of type ****. There are about 30 of these.
NumericValue	The root of the numeric value types.
ReferenceValue	The root of the value classes which reference another value.
InvariantValue	A ReferenceValue which includes an invariant function.
UpdatableValue	A ReferenceValue in which the referenced value can be changed.
NameValuePair	A class to hold a name and a runtime value pair.
NameValuePairList	A list of name/value pairs.
NameValuePairMap	A map of name/values.
Quantifier	A class representing a quantifier.
QuantifierSet	A class representing a set of quantifiers.
State	A class for holding a module's state data.
ValueList	A sequential list of values.
ValueMap	A map of value/values.
ValueSet	A set of values.

The vdmj.values class contains a set of value classes to represent runtime values in the interpretation of a specification. They are all subclasses of the abstract Value class.

All Values implement a set of standard methods to allow them to work correctly with the Java collections framework: equals, hashCode, toString, clone.

The convertTo method takes a Type argument and is responsible for the dynamic type conversion of values from one type to another, where possible. This tests the -dtc settings flag, returning the value unchanged if the flag is set. The method is used when types are “enforced” in a specification, such as when values are passed as typed arguments, or when values are returned from a typed function or operation. The method is specialized in all the Value subclasses that generally know how to convert themselves into a small number of related types (eg. conversions between subclasses of NumericValue). If a value cannot convert itself, it delegates to its superclass. The convertTo at the root of the Value hierarchy deals with common complex cases: converting to a union (iteratively trying to convert to each type); converting to a parameter type (looking up the parameter’s name to get its actual type); converting to optional types (convert to the underlying type); and converting to a named type (convert to the underlying type, then wrap with an InvariantValue to enforce any type invariant).

A ReferenceValue is an abstract class that contains a value, and delegates all operations to that contained value. The concrete subclasses are InvariantValue, which includes an invariant FunctionValue as well as the value; and UpdatableValue, which has a set method capable of changing the value referenced (and all its methods are synchronized to allow for safe access to the changed value from other threads). When an UpdatableValue is created, it is passed a ValueListener (optionally), which is called back when the set method is called. This is used to invoke class or state invariant functions when values are changed.

Most value classes are immutable, in the sense that they are constructed with an underlying value that is held in a (public) final field of the class, and never changes. The only exception to this is the UpdatableValue, whose referenced value can be modified by the set method – note this changes the immutable value referenced; it does not change the value of the referenced object.

UpdatableValues are used for “state” values (instance variables, module state and “dcl” values that can be changed in an assignment). They are often initialized with structured immutable values, so values implement a method called "getUpdatable" to create UpdatableValue versions of themselves and their sub-values.

A Value can be held together with a LexNameToken in a NameValuePair, and such pairs can be collected into a list or a map for convenience. For example, the members of an ObjectValue are held in a NameValuePairMap.

Internally, many value classes have basic Java types at their heart. The basic values underlying SetValues, SeqValues and MapValues are ValueSets, ValueLists, and ValueMaps respectively. The basic value of a Value can be extracted using one of several methods, like realValue or functionValue. If the Value concerned cannot be converted to the basic type sought, a ValueException is thrown. The basic underlying values are used in the eval method of expressions to (say) actually perform arithmetic.

2.16.1. Comments

It seemed sensible to have all Values contain some sort of primitive value, including sets, sequences and maps, rather than having the Value classes extend the Java collection framework directly. That allows the Value types to be in a "VDM" hierarchy, but it leads to the confusing situation where a MapValue contains a ValueMap – the former extends Value, the latter extends HashMap<Value,Value>. The ValueXXX classes should only be used internally – ie. they should never be returned from an eval method (they can't be, as they're not Value subtypes).

2.17. vdmj.commands

Class Summary	
CommandReader	A class to read and perform commands from standard input.
ClassCommandReader	A class to read and perform class related commands from standard input.
ModuleCommandReader	A class to read and perform module related commands from standard input.
DebuggerReader	A class to read and perform debugging commands from standard input.

The vdmj.commands package contains the command line readers that implement the interactive actions of the suite. This should be the only package that interacts with the user terminal.

A subclass of CommandReader is instantiated to provide the main command prompt of the interpreter: either ClassCommandReader or ModuleCommandReader. These classes can set or clear breakpoints, but they do not permit commands like “step” and “next”, as these are only legal from a breakpoint context. The (few) differences between the two classes concern the commands that are legal for each – eg. "classes" and "modules" – while the truly common code is in the parent.

At a breakpoint, a DebuggerReader is used to permit the extra debugging commands allowed from this context, and prevent the ones that aren’t (like initializing the environment on the fly). The one DebuggerReader is used to handle debugging sessions for VDM-SL and VDM++.

2.17.1. Comments

The idea here is to isolate the console interaction from the rest of the program so that the core interpreter could be included in (say) an Eclipse plugin. In practice, both Tracepoints and Stoppoints have to write to the console as well. This area needs some work to completely separate the implementation from the console control, so that breakpoints could interact with (say) Eclipse.

2.18. vdmj.messages

Class Summary	
VDMMessage	The root of all reported messages.
VDMError	A VDM error message.
VDMWarning	A VDM warning message.
Console	A class to provide System.in/out with a charset encoded wrapper.
Redirector	An abstract class to redirect output.
StdoutRedirector	A class to redirect standard output to a debugger.
StderrRedirector	A class to redirect standard error to a debugger.

Exception Summary	
MessageException	An exception to carry a textural error message.
NumberedException	An exception to carry a number and text information.
LocatedException	An exception to carry number, text and location information.

The vdmj.messages package contains classes and exceptions to hold and display error and warning messages. Lists of VDMMessage values are returned from the TypeChecker and Reader classes. Similarly, internal exceptions that carry message numbers and position information are all subclasses of NumberedException.

The Console class manages output to stdout/stderr, in particular managing the character set to be used. The Redirector and its concrete subclasses are used to redirect or copy stdout and stderr to a debugger (see 2.18).

2.19. vdmj.debug

Class Summary	
DBGPReader	The main class of the DBGp protocol reader.
DBGPCommand	A parsed IDE command.
DBGPOption	A parsed IDE command option.
DBGPFeatures	The set of features supported/set by the debugger.

Enum Summary	
DBGPBreakpointType	The possible DBGp breakpoint types.
DBGPCommandType	The possible DBGp IDE command types.
DBGPOptionType	The possible IDE command option flags.
DBGPErrorCode	The possible status response error codes.
DBGPReason	The possible status response reason codes.
DBGPContextType	The possible variable name context types.
DBGPRedirect	The possible I/O redirection options.
DBGPStatus	The possible status responses.

Exception Summary	
DBGPException	A general purpose debugger exception.

The vdmj.debug package contains classes that implement the DBGp remote debugging protocol defined in [6]. This allows VDMJ to load a set of specifications and evaluate/debug them under the remote control of another process – usually a GUI IDE, such as Eclipse.

As described in [6], the principle behind DBGp is that the IDE launches the debugged process (ie. VDMJ in our case), and that process connects back to the IDE using a TCP/IP connection to a host/port specified by the IDE. The IDE then sends commands to the debugger on the connection, and the debugger acts on those commands, sending status and results back to the IDE via the same connection.

Because DBGp starts VDMJ, the principal class that handles the connection, DBGPReader, defines a "main" method. This is separate from the main method defined in the VDMJ class (see 2.1). The command line arguments are as follows:

```
<host> <port> <ide key> <dialect> <expression> {<filenames>}
```

The host and port identify the connection that must be opened back to the IDE to receive commands. The ide key is a value that must be passed back to the IDE during the initial connection. The dialect is one of the values of the Dialect enum ("VDM_SL" or "VDM_PP", see 2.2), the expression is the main expression to be evaluated, and the list of filenames identify the specification itself. Each filename is in the form of a file URI ("file:/...").

Depending on the dialect, DBGPReader creates an instance of VDMPP or VDMSL (see 2.1), and uses it to parse and type check the list of files passed. The DBGp protocol assumes the "program" being debugged is already compiled correctly, so if there are any syntax or type checking errors, these are sent to the VDMJ process' standard output and the process quits with an error exit code. The protocol has no mechanism for returning these errors to the IDE.

If the specification is clean (warnings are permitted), an Interpreter object encapsulating the parsed/checked specification is obtained from the VDMPP or VDMSL object, and this, along with the host, port, ide key and expression are passed to the constructor of a new DBGPReader object. The constructor saves the values, and sends the DBGp initialization message back to the IDE. Lastly, the main method calls the run method of the reader. When this method returns, the VDMJ process quits with a success error code (0). If any exceptions are thrown from run, it quits with an error exit code.

The DBGPReader run method is a read/execute loop, reading DBGp commands from the IDE connection, processing them, sending back any responses, and then looping. The return value from the private methods to handle each type of command indicate whether the run method should keep looping or return (for example, the "detach" command would indicate that the loop could return, though most commands keep looping).

All DBGp commands are in a simple textural format, similar to the format of a UNIX command (see [6]):

```
command -op1 v1 -op2 v2 ... -- data
```

This text is parsed by a method which produces a DBGPCommand object; parsing errors throw a DBGPException, which is caught and used to build an error status response before returning to the run loop.

Successfully parsed commands are passed to one of a number of processCmd methods which handle each type of command. The general form of these is to validate the options passed (held in the parsed DBGPCommand object), and send back an error response if not correct; and then act on the command, sending back a successful status, possibly combined with information being requested by the IDE.

Responses sent to the IDE are all XML formatted messages (see [6]). A collection of private methods

in DBGPReader take the raw text response from a processed command in a StringBuilder, and use this to create an XML response message and send it on the connection to the IDE .

All of the DBGp commands which interact with the running specification do so by making method calls on the Interpreter object passed to the DBGPReader's constructor. Note that this is the abstract Interpreter class, not the concrete ClassInterpreter or ModuleInterpreter, so the DBGPReader will work with VDM-SL or VDM++.

When the specification is executed (via the DBGp "run" command), the interpreter's execute method is called, passing the expression to be evaluated. The DBGPReader instance is also passed to the execute method; this is stored in the thread context information for the main VDM thread, and allows breakpoints to find the connection to the IDE.

The interpreter's execute method does not return until the specification has been fully evaluated, so this raises the question of what happens at breakpoints, when information must be passed to the IDE and more instructions received.

Breakpoints normally use the DebuggerReader class to interact with the user on the command line before continuing execution (see 2.14). But when the process is being remotely debugged, the breakpoint will find the DBGPReader object associated with its thread context, and call its "stopped" method. This sets the state of the debugger to "break" and sends a status message to the IDE to let it know that execution has stopped at a breakpoint. It then enters the "run" method to process IDE commands as it did originally (in fact this is a recursive call, since the original run call is on the stack, having called the execute method). Note that some DBGp commands are only acceptable when the debugger is in the "break" state, such as "step_into" and "stack_get". Expressions can be evaluated in the "break" state and will be evaluated in the context of the breakpoint (ie. local variables are visible).

A "continue" command from the IDE will cause the recursive run call to return, and the flow of control will return to the breakpoint and from there back into the main execution. Further breaks may occur, but eventually, the main evaluation will complete and the original run method call will regain control. Note that the debugged process does not automatically quit at this point. It enters a "stopped" state, where further (restricted) IDE commands are possible, such as to retrieve the final evaluation result.

The DBGp protocol allows multi-threaded programs to be debugged. In this case, each thread opens its own connection to the IDE (on the same host/port). The IDE is responsible for sending commands for each thread on the appropriate connection. As far as the debugged process is concerned, these connections are completely separate. Therefore, when VDM++ starts a new thread, the creating thread's DBGPReader is "cloned" to create a new object which will open its own connection to the IDE, and be stored in the new thread's context. When the second thread reaches a breakpoint, it will respond to the IDE on its own connection. The main thread may still be running – if the IDE chooses to stop all threads at a breakpoint, it must do so by interrupting the other threads by sending commands on their connections, though to do so, the DBGPReader must be able to handle asynchronous commands (receiving commands while it is running, not just at breakpoints). This functionality is not yet available.

The IDE can request that stdout/stderr output from the debugged process be sent to the IDE rather than (or as well as) to the usual console. If such a command is received, the Console class is called (see 2.17), and a Redirector is added to the appropriate stream. These objects are passed a DBGPReader reference and use this to send output to the IDE when directed to do so.

2.19.1. Comments

It would probably be better to use some sort of XML handling package to build the IDE responses, rather than hand-crafting them. Though the XML involved is very simple.

Asynchronous debugging could be done simply if checkpoints (ie. at the start of every expression or statement) check the status of the DBGp connection. That is simple to arrange, as the object is in the thread's context, but the overhead may be large (calling the input stream's "available" method). It might be acceptable to (say) only test the stream every 100 or 1000 operations, at the risk of making the threads unresponsive to asynchronous breaks.

A better solution would be to have asynchronous listening threads that call the "interrupt" method of VDM threads.

If you attempt to evaluate something at a breakpoint which causes another breakpoint to be hit, the system will trap back into the DBGPReader again (ie, three re-entrant calls to "run"). This should behave as you expect (ie. continue will take you back out to the first breakpoint), but whether it does what the Eclipse IDE expects remains to be seen. Ideally, breakpoints need to be disabled while stopped at a breakpoint, though there may be issues with multiple threads which can still be running.

Note that the Console class only has one Redirector wrapped around an output stream. That means that all output is redirected to one particular DBGp connection, rather than the output from different threads being directed to each thread's DBGp connection. The thread that receives all the output is the one that last send the IDE command to redirect it.

Currently, the values of all variables are sent back to the IDE as strings. To enable the structure of aggregate types to be expanded by the IDE, we need to define an XML Schema definition to describe them. This has not yet been done. When it is defined, it should be possible to change the Type class hierarchy to produce XML Schema to describe themselves, and to change the Value class hierarchy to "print" themselves in that schema.

2.20. vdmj.util

Class Summary	
Utils	A utility class.
IO	The native IO library
VDMUtils	The native VDMUtils library
MATH	The native MATH library
Base64	A class to encode/decode base64 strings.

Utils is a small utility class. It just defines a few methods for printing out a comma separated list of arbitrary types held in a Java List<T>.

The IO, VDMUtils and MATH classes implement the native interceptors for the standard CSK library routines. These are “not yet implemented” in the VDM source code, but the VDMJ classes for “not yet implemented” will intercept various reserved function/operation names and call these class' static methods instead.

Base64 does what you would expect. It is used by the DBGp protocol.

