The design and implementation of a userfriendly object-oriented markup language

Titouan Vervack

Supervisor: Prof. Marko van Dooren

Counsellor: Benoit Desouter

Master's dissertation submitted in order to obtain the academic degree of

Master of Science in Computer Science Engineering

Vakgroep Toegepaste Wiskunde, Informatica en Statistiek Chair: Prof. dr. Willy Govaerts Faculty of Engineering and Architecture Academic year 2015-2016



The design and implementation of a userfriendly object-oriented markup language

Titouan Vervack

Supervisor: Prof. Marko van Dooren

Counsellor: Benoit Desouter

Master's dissertation submitted in order to obtain the academic degree of

Master of Science in Computer Science Engineering

Vakgroep Toegepaste Wiskunde, Informatica en Statistiek Chair: Prof. dr. Willy Govaerts Faculty of Engineering and Architecture Academic year 2015-2016



#### i

## Preface and acknowledgement

In the first place I thank my promotor, professor Marko van Dooren, for all of the guidance, encouragement and feedback he provided to me during the completion of this thesis. The thesis would not have reached the current state without our weekly meetings.

My thanks also go out to my counselor, Benoit Desouter, for proofreading the final work on such short notice.

I also thank Stijn Seghers for providing feedback on my work and for listening to my rambling about said work.

Lastly, I thank my family and friends for their support during this year. In particular I thank my close family, Marino, Nathalie, Valentin and Amaury, for all the encouragement and support they have given me throughout my studies.

Titouan Vervack, June 2016

## Permission for usage

"The author(s) gives (give) permission to make this master dissertation available for consultation and to copy parts of this master dissertation for personal use. In the case of any other use, the copyright terms have to be respected, in particular with regard to the obligation to state expressly the source when quoting results from this master dissertation."

Titouan Vervack, June 2016

# The design and implementation of a userfriendly object-oriented markup language

by

#### Titouan VERVACK

Master's dissertation submitted in order to obtain the academic degree of Master of Science in Computer Science Engineering

Academic year 2015–2016

Supervisor: Prof. dr. ir. Marko van Dooren Counsellor: Dr. ir. Benoit Desouter

Faculty of Engineering and Architecture, Ghent University

Department of Applied mathematics, computer science and statistics, Chair: Prof. dr. Willy Govaerts

#### Abstract

A lot of solutions exist for document creation. These are split up into What You See Is What You Get (WYSIWYG) editors and markup languages. Both of these provide a lot of advantages as well as disadvantages.

WYSIWYG editors immediately show the final document and help the user by providing a GUI for all the features. Markup languages put greater emphasis on letting the user structure a document properly, by providing specialised syntax for certain structural elements for example. Because of this some solutions are easy to read and write while others offer a lot of customisability, usually through some programming model. However, a good combination of both is hard to find.

As such the goal of this thesis is to design a markup language that is easy-to-read and write as well as providing a modern programming model to allow for full customisation of the document. An introduction to the state-of-the-art solutions for document production is given (Chapter 1) in which the (dis)advantages of said solutions are discussed. This is followed by an introduction to the new markup language, its features and the design decisions that were made (Chapter 2). Then several proofs of concept for document types and libraries build with the language are presented (Chapter 3). The development of the compiler for the language, its features and the faced challenges are then discussed (Chapter 4). The text concludes with possible future work on this work (Chapter 5), as well as a short reflection and summary on the work (Chapter 6).

## Keywords

Markup, Object-oriented, LaTeX, Markdown

*CONTENTS* v

## Contents

P	Preface and acknowledgement									
P	ermis	ssion for usage	ii							
1	Inti	roduction	1							
	1.1	State-of-the-art	1							
		1.1.1 Word processors	3							
		1.1.2 IATEX	5							
		1.1.3 Markdown	8							
		1.1.4 Pandoc	9							
	1.2	Problem statement	11							
		1.2.1 Word processors	12							
		1.2.2 IATEX	13							
		1.2.3 Markdown	14							
		1.2.4 Pandoc	15							
	1.3	Proposed solution: Neio	16							
		1.3.1 Target group	17							
<b>2</b>	Des	sign of the Neio markup language	18							
	2.1	Neio document	18							
	2.2	Code files	19							
	2.3	Call chain	21							
	2.4	Method modifications	23							
		2.4.1 Symbol methods	23							
		2.4.2 Text	23							
		2.4.3 Surround methods	24							

*CONTENTS* vi

		2.4.4	Nested methods	25
		2.4.5	Newlines	27
	2.5	Conte	xt types	31
	2.6	Code	blocks	42
		2.6.1	Non-scoped code	43
		2.6.2	This	44
		2.6.3	Scoped code	45
		2.6.4	Inline code	46
	2.7	Text is	n code mode	47
	2.8	Naviga	ating through the lexical structure	48
	2.9	Consid	derations	50
		2.9.1	Static typing	50
		2.9.2	Security	50
3	Sun	ported	l document types and libraries	<b>52</b>
J	3.1	_	nent classes	52
	3.1	3.1.1	Document class definition	$\frac{52}{52}$
		3.1.2	Document	53
		3.1.3	Book	55 55
		3.1.4	Abstract	56
	3.2	_	ies	58
	0.2	3.2.1	BibTeX	58
		0	References	60
		3.2.3	LATEX math and amsmath	61
			ETEX math and amsmath	
		3.2.4 3.2.5	L	63
			Red black trees	66
		3.2.6	MetaUML	69 70
		3.2.7	Chemfig	72
		3.2.8	Lilypond	74
4	Imp	lemen	tation	77
	4.1	Compi	ile flow	77

*CONTENTS* vii

	4.2	Used libraries and frameworks	80
		4.2.1 ANTLR4	81
		4.2.2 Chameleon and Jnome	81
	4.3	Translation	82
		4.3.1 Reasons for choosing Java	82
		4.3.2 Fluent interface	83
		4.3.3 Outputting	83
		4.3.4 Reflection	85
		4.3.5 Escaping	88
	4.4	Resource usage and creation	89
	4.5	Limitations	89
		4.5.1 Windows	89
		4.5.2 Back end	89
_	T- 4		01
5	ruti	ure work	91
	5.1	Automatisation	91
	5.2	Moving content	93
	5.3	Use	94
	5.4	Packages	94
	5.5	Compiler improvement	94
	5.6	Tool improvement	95
6	Con	aclusion and reflection	97
J	COI		01

LIST OF FIGURES viii

## List of Figures

1.1	The detection of an error	2
1.2	The warning before trying to repair the error	2
1.3	The inability to repair the error	3
1.4	An Excel chart in a Word document	4
1.5	Kerning on	7
1.6	Kerning off	7
1.7	Ligature on	7
1.8	Ligature off	7
1.9	The document rendered as HTML by the official Markdown conversion tool	9
1.10	The output of the document to the left	0
1.11	The output of the document to the left	1
1.12	Caption below the image	3
1.13	The image moved, but the caption did not	3
1.14	The table document rendered as HTML by the official Markdown conversion	
	tool	5
2.1	The rendered form of the document to the left	9
2.2	The object model of the previous document	2
2.3	The rendered version of the document to the left	5
2.4	The tree representation of the call chain of the previous example $3$	2
2.5	The first context type: ct0	3
2.6	The context types ct0 and ct1	4
2.7	The context types ct0 through ct2	5
2.8	The context types ct0 through ct5	6
2.9	The context types ct0 through ct6	7

LIST OF FIGURES ix

2.10	The object model of the previous example	39
2.11	Correct context lookup for *	39
2.12	Wrong context lookup for *	40
2.13	The correct context type representation	41
2.14	The wrong context type representation	42
2.15	The rendered form of the example to the left	47
2.16	The rendered form of the example to the left	48
2.17	The rendered version of the example to the left	49
3.1	The rendered form of the example to the left	62
4.1	Illustration of how a Neio document is compiled. Neio code files translate to	
	Java classes, Neio script files to a Java file with a main method. This then gets	
	output to TeX using the Neio standard library	78

*LISTINGS* ×

# Listings

1.1	document.md	9
1.2	pandocExample.md	10
1.3	behead2.hs	10
1.4	letterTemplate.latex	11
1.5	letter.md	11
2.1	A simple Neio document	19
2.2	Paragraph.no	19
2.3	Document.no	21
2.4	Text.no	24
2.5	surroundEx.input	25
2.6	Chapter.no	26
2.7	TextContainer.no	26
2.8	Two lists	27
2.9	One list	27
2.10	TextContainer.no	28
2.11	NLHandler.no	28
2.12	ParNLHandler.no	29
2.13	ParNLHandler.no	29
2.14	A table created using a DSL	30
2.15	An example for context resets	38
2.16	Automatic return from code block	44
2.17	Manual return from code block	44
2.18	template.no	47

*LISTINGS* xi

2.19	dice.no	48
2.20	Content.no	49
3.1	Content.no	53
3.2	Thesis.no	55
3.3	thesis.no	56
3.4	overview.no	57
3.5	AbstractNLHandler.no	57
3.6	Bibtex.no	59
3.7	Document.no	59
3.8	Chapter.no	61
3.9	Referable.no	61
3.10	An ASCII table created in Neio.	63
3.11	TextContainer.no	64
3.12	P TableRow.no	65
3.13	TableNLHandler.no	65
3.14	l rbtDocument.no	66
3.15	6 rbtLatex.tex	68
3.16	StructureNLHandler.no	73
3.17	7 Atom.no	73
3.18	3 An example musical score	74
	9 Score.no	74
	) Score.no	76
3.21	The code of the example.	76
4.1	TexFileWriter.no	79
4.2	TexToPDFBuilder.no	79
4.3	project.xml	82
4.4	testInput.no	84
4.5	testInput.java	84
4.6	Content.no	86
4.7	Text.no	88
5.1	Content.no	91

LISTINGS	xii

5.2	InlineEq.no	 				•				 			 				92	

INTRODUCTION 1

### Chapter 1

## Introduction

In this thesis we introduce a new markup language that improves upon a few others, namely LATEX and Markdown, whilst still holding on to their advantages.

The language has been used to write this entire book with the exclusion of the title page and a few images. The source code of the book is available on the UGent GitHub [22]. In the same repository you also find the Neio library [21] and source code for the Neio compiler [31].

Before we get into the details of this new language it is necessary to introduce some of the most used markup languages and word processors.

#### 1.1 State-of-the-art

Before we discuss the state-of-the-art, we first discuss the difference between saving a document in a human-readable or a binary format.

Markdown, LATEX and Pandoc use a human-readable format that can be opened in any editor, is quite robust to file corruptions and is well suited for Version Control Systems (VCS) such as Git and Mercury. It is robust in the sense that a if a single bit of the document gets flipped, the document can still be repaired by the user. The user can find the error on his own, or he can use the variety of editors to find help him as every editor can handle errors differently.

When the syntax gets more extensive and less human-readable, the chances to encounter a problem upon compilation increase. This is both good and bad, if a part of the syntax got corrupted, the compiler tells you and you know that the document has been corrupted. Had it not warned the user, the document would be in a corrupt state without the user noticing so. On the other hand the corruption can prevent the document from successfully compiling, forcing the user to find out what happened to the document. In any case however, the error can still be tracked down and repaired.

This is not always possible, when using a binary format, like the ones used by default in What You See Is What You Get (WYSIWYG) editors for example. As a matter of fact, we created a Word document and flipped one bit in a random byte using a hex editor. We then tried to open the document and received the following sequence of errors and were unable to retrieve the contents of the file.

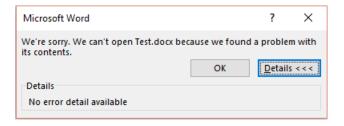


Figure 1.1: The detection of an error.

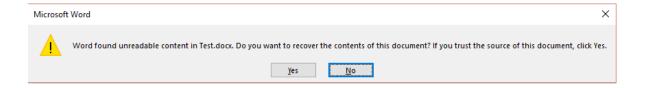


Figure 1.2: The warning before trying to repair the error.

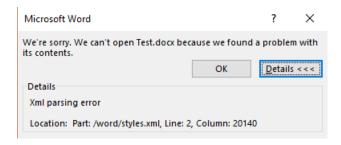


Figure 1.3: The inability to repair the error.

A regular user, without technical expertise, is not be able to repair this error as the source code is not human-readable. There are also almost no other tools to help the user as the binary format can only be interpreted correctly by a few editors.

It is to be noted that the error handling depends on what exactly was corrupted, in some cases the corruption can be automatically repaired.

#### 1.1.1 Word processors

The best-known solution for creating documents, are so called WYSIWYG editors, examples include 'Microsoft Word' and 'Pages'. They are called as such because you immediately, without compiling, see what your final document is going to look like. Because of this, the complexity of the tool is significantly decreased allowing anyone to use the tool. It is less complex because you do not have to type in commands and wait for a preview to change. Instead you select the property from a drop-down list or click a button and the document is updated instantly. The GUI essentially hides the complexity.

A WYSIWYG editor gives the author freedom, for example it is very easy to change a font or change the font size of a certain part of the text and you can easily add an image and drag-and-drop it around.

A lot of these editors are grouped into packets, such as Office or LibreOffice. Using one of these packets, it is also easy to insert slides, tables or another document [23]. These are then be editable with the specialized program for it (like Excel), but you do not have to leave the

document. In the example below an Excel chart has been inserted into a Word document.

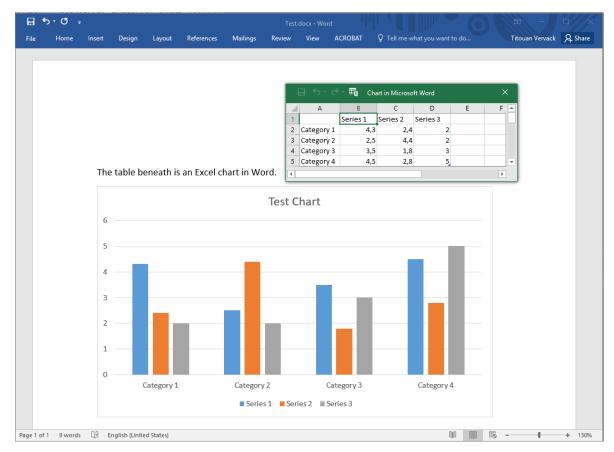


Figure 1.4: An Excel chart in a Word document.

Collaboration is also made easy by using online versions of these solutions, such as 'Microsoft Word Online' or 'Google Docs'. These make documents cross-platform and allow multiple people to work on them at the same time. It is also possible to add comments or suggestions on (online) documents. These changes can then be resolved in just one click by an authorized editor.

Some documents are however more complex and have need for a more powerful tool such as  $\LaTeX$ .

#### 1.1.2 LATEX

LATEX is a document preparation system that uses the TeX typesetting system. It provides a set of extra macros that can be used on top of the TeX macros.

LATEX provides rich customisability through the Turing complete programming model provided by TeX. The model is different from modern programming models, it does not use regular variables or functions, instead it uses macros. A macro or command looks like the following: \cmdname[opt-arg]{req-arg}. Here cmdname is the name of the command, opt-arg are the optional arguments and req-arg are the required arguments. To show how commands work, we show the process involved in processing some input that includes a command.

In LATEX, every character and command in the input is a token, a is a token but \command is also a token. The input is transformed into a list of tokens and is then processed token after token. When a command is encountered, it is expanded.

Expansion removes the current token from the list and adds every expandable item in its body to the token list recursively. This is best shown using an example.

We define two commands using \newcommand and call the \y command.

```
1 \newcommand{\x}{a command}
2 \newcommand{\y}{This is \x}
3 \y
```

The output is This is a command. The step-by-step expansion is given below:

- 1. LATEX encounters the token \y without arguments and expands it.
- 2.  $\forall y$  is removed from the token list and This is  $\forall x$  is added to it.
- 3. This is is read and  $\x$  is encountered,  $\x$  gets expanded.
- 4. \x is removed from the token list and a command is added.

We can also pass arguments to a command by specifying the amount of arguments as an optional argument. We can then access the arguments by using the hash symbol followed by a number. For example, the output of code beneath is: This is an argument.

```
\newcommand{\command}[1]{This is #1}
command{an argument}
```

The braces around the first required argument are optional, thus we can also write this as follows:

```
\newcommand\command[1]{This is #1}
```

To elaborate further on expansion, we have a look at another example. We have the string Hello saved in a command and want to concatenate a string to it so that it prints out Hello world. To do this, we have to redefine the command, this can be done using \renewcommand.

We thus get the following:

```
1 \newcommand\example{Hello}
2 \renewcommand\example{\example{} world}
```

The {} at the end of \example is used as a delimiter, to know when the \example command ends, such that the space behind it does not disappear. This example does not work however as \example is recursively calling the new definition instead of calling the old one once. To counter this, \example in the new definition should be expanded first, we can to that using \expandafter as shown below.

```
\newcommand{\example}{Hello}
converge \text{\example} \expandafter\renewcommand\expandafter\example\expandafter{\example} \world}
converge \text{\example}
```

\expandafter removes the first two tokens behind itself from the token list, expands the second token and then adds the expanded token and the first token back to the token list.

The step-by-step expansion for the previous example is given below:

- 1. \example is defined for the first time.
- 2. The second definition begins.
- 3. \expandafter removes \renewcommand and \expandafter and expands the latter.
- 4. The second \expandafter removes \example and \expandafter and expands the latter.
- 5. The third \expandafter removes { and \example and expands the latter to Hello.

- 6. \renewcommand, \example, \{\) and Hello are placed back on the token list.
- 7. The token list (\renewcommandexample{Hello world}) is now expanded normally.

Using this programming model, libraries, called packages can be created. Over the course of the past 31 years (the initial LATEX release was in 1985), a lot of them have been created. They offer all kinds of functionalities, ranging from the creation of sheet music and slide shows [28] to providing the base to write letters and books. This is one of the reasons why LATEX is often used for long and complex documents such as books, scientific papers and syllabi.

The goal of LaTeX [13] is for the user to focus on the structure and content of the document and to let designers worry about the design of the document. To make sure the document looks good afterwards, it is important to structure it using elements such as sections and paragraphs. Because of the freedom given in WYSIWYG word processor, it is often forgotten to use styles, headers and so on. This can cause inconsistencies to appear as the document grows.

Designers can use the very sophisticated typesetting of LATEX to create designs for well structured LATEX documents. The typesetting improves readability by supporting kerning, ligatures,... and using the advanced Knuth-Plass line-breaking algorithm [11]. The algorithm sees a paragraph as a whole instead of using a more naive approach and seeing each line individually. Kerning places letters closer together or further away depending on character combinations. A ligature is when multiple characters are joined into one glyph.



Figure 1.5: Kerning on Figure 1.6: Kerning off

fi fi

Figure 1.7: Ligature on Figure 1.8: Ligature off

LATEX also knows how to work with bibliographies by using programs such as Bib(La)TeX or Biber. These are reference management programs that are very robust and make certain

that all of your references are consistent. They allow you to create the references outside of your main document, decreasing the amount of clutter in the document.

Whilst on the topic of including other files in your document, in LATEX you can split up your document into multiple smaller documents, allowing for easier management. A common use of this function is to write every chapter, for example in a book, in a separate document and than all the chapters are imported into the main document. This allows to easily remove or switch out different parts of a document. This is also possible in Pandoc (as you can inline LATEX) and the WYSIWYG editors [23].

Finally, IATEX employs a free and open source model, which is a big reason why it is so popular in the open source community. As a result of this model, combined with its popularity and good scientific support, LaTeX has been integrated in many other applications. A few examples are:

- MediaWiki [17] [30]: a free and open-source software package that powers sites as Wikipedia [34] and Wiktionary [35].
- Stack Exchange [33]: a network of Q&A web sites
- JMathTex [9]: a Java library adds functionality to display mathematical formulas in a Java application

IATEX is powerful but it is also complex and has a steep learning curve, because of this, simpler alternatives such as Markdown have been created.

#### 1.1.3 Markdown

The goal of Markdown [15] is to be easy-to-write and read as a plain text document and as such is actually based on plain text emails. Its appeal lies in its simplicity as it allows to for example easily create a simple document, such as a short report or a blog post.

It achieves this simplicity, by introducing a small and simple syntax [16] that feels natural. For example, to create a title you underline it with minus or equals characters. To create an

enumeration, you use star characters as bullet points. An example Markdown document is shown below:

## Listing 1.1: document.md

```
Markdown

=======

It is:

* Simple

Easy to read

* Easy to write
```

- It is:
- Simple

Markdown

- Easy to read
- Easy to write

Figure 1.9: The document rendered as HTML by the official Markdown conversion tool.

This syntax decreases the amount of possible syntax errors and allows for short compile times and wide editor support. It also allows Markdown to be easily translated into other formats such as PDF or HTML. As such, a tool to convert Markdown to HTML is offered by the designers of the language itself. This tool is also cross-platform and as such widens the target audience of Markdown.

However, for a lot of cases Markdown is too simple. Because of this, solutions such as Pandoc have been created.

#### 1.1.4 Pandoc

Pandoc [24] is a document converter, but it can do much more than just convert a document from one format to another. Notably, it allows you to write inline LATEX in a Markdown document, combining the power of both. An example is shown below.

Listing 1.2: pandocExample.md

#### Pandoc

This pandoc document uses LATEX and markdown in one file!

$$\sqrt[3]{125} = 25$$

Figure 1.10: The output of the document to the left.

However, outside of the LATEX one, Pandoc does not really have a programming model and instead works with so called filters. A document in Pandoc can be read and transformed into an Abstract Syntax Tree (AST). The AST is read in a filter, then transformed and passed on to the next filter.

Filters can be written in a multitude of languages such as Haskell, Python, Perl,..... They are then passed as a command line argument to the conversion command. We illustrate this using an example from the Pandoc tutorial on scripting [25].

Listing 1.3: behead2.hs

```
#!/usr/bin/env runhaskell
-- behead2.hs
import Text.Pandoc.JSON

main :: IO ()
main = toJSONFilter behead
where behead (Header n _ xs) | n >= 2 = Para [Emph xs]
behead x = x
```

(Pandoc, 2016)

The example above shows a filter written in Haskell. It replaces all the headers of level 2 or more in a Markdown file by paragraphs with the text in italics (hence the Emph in the code). The toJSONfilter function constructs a JSON representation of the AST. Using this, we find the headers of level of 2 or more and create a new paragraph. The document can be created using the following command pandoc -f SOURCEFORMAT -t TARGETFORMAT --filter ./behead2.hs.

This example shows the power of Pandoc: it can do its work directly on the AST. This means

that the source document does not have to be a Markdown document, it could also use HTML or any other supported input format.

Using another one of these filters, PanPipe [32], you can execute programs defined outside of a document. The stdout of the program is inserted into the document and the stderr is redirected to the stderr of Pandoc. An example is shown below.

```
This document executes a shell command.

'''{pipe="sh"}

echo "Hello world! I am " $(whoami)

'''
```

This documents executes a shell command. Hello world! I am titouan

Figure 1.11: The output of the document to the left.

Using Pandoc, you can also declare templates that allow you to use variables. The value of the variables is predefined (such as \$body\$) or passed in through the command line. These can then also be used in conditionals or loops in this template. You can see an example below.

Listing 1.4: letterTemplate.latex

Listing 1.5: letter.md

```
1 \documentclass{letter}
2 \signature{Titouan Vervack}
3 \begin{document}
4 \opening{Dear $addressee$,}
3
5
6 $body$
7
8 \closing{Sincerely,}
9 \end{letter}
10 \end{document}
```

```
How have you been?
I've been longing to see you, but I've been very busy...

My son's getting married next week and I would like to invite you!
Well, you and your wife of course.

If you're planning on coming, please give me a call.
```

The document can then be build with the following command.

```
pandoc letter.md -o letter.pdf --template letterTemplate --variable=addressee :"John Doe"
```

#### 1.2 Problem statement

We can see that these solutions all have their advantages, but there are also quite a few disadvantages. We discuss the most common ones for every solution underneath and afterwards

we introduce the problem statement.

#### 1.2.1 Word processors

The UI in WYSIWYG editors allows you to customise the document, you can insert other documents, tables, figures,... and adjust a lot of properties. It is also possible to use macros using for example Visual Basic for Applications in Word to automate some tasks. It can for example be used to remove leading tab characters from paragraphs.

However, it has trouble with inserting variables. Because of this, it is very cumbersome to compute a simple expression like x + y. To do this, you first have to open the Visual Basic Editor. In there you create a new module in which you define your new variables as follows.

```
ActiveDocument.Variables.Add Name:="x", Value:=27
ActiveDocument.Variables.Add Name:="y", Value:=2
ActiveDocument.Variables.Add Name:="result", Value:=(Int(ActiveDocument. Variables("x")) + Int(ActiveDocument.Variables("y")))
```

The result variable can then be inserted by selecting Quick parts in the Insert tab, selecting Field, then selecting DocVariable and finally typing in result. This inserts the result variable into the document, but the variable is not automatically update. To achieve that you have to right-click it and click Update field.

In word processors small changes can also have big side effects on the document. For example, you placed an image exactly where you wanted and later on, you add a newline somewhere higher up. The reflowing of the text underneath it shifts the precisely placed image out of place. These changes are not always immediately clear, which creates inconsistencies in the document.

Another problem that occurs with images (thought it is not an inherent problem of WYSI-WYG editors, rather a bad implementation), is that the caption does not move together with the image. In the example below an image is inserted and a caption is (by right-clicking the image and pressing Add caption) added. The image is then dragged to the end of the file and we see that the caption remains in place.

This is the first paragraph.

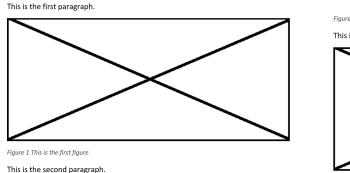


Figure 1.12: Caption below the image.

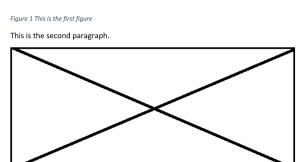


Figure 1.13: The image moved, but the caption did not.

#### 1.2.2 LATEX

Even though LaTeX offers customisability through the programming model of TeX, it is not perfect. The programming model is hard to use and read, creating a steep learning curve. It also makes LaTeX seem complex, scaring away a lot of potential users. A way to counter this complex look, is LyX [14]. This is a GUI around LaTeX offering some of the most common WYSIWYG features, while still utilizing the strength of LaTeX. Even though the complexity can be hidden under a GUI, it is still there and error messages are as bad as they were before, as is discussed beneath.

Having a cumbersome programming model, TeX makes it easy to create syntax errors. When these are made TeX shows an error message, but these are often unclear. See the excerpt below for an example.

```
line 4: Display math should end with $$

\documentclass{article}
\begin{document}
\begin{equation}
\document
\docume
```

The message means that \$ is not allowed inside of an equation, but that is not clear from the

error message.

As mentioned in Subsection 1.1.2, the macro model does not work as we are used to in imperative and functional languages, we have to think about the order of expansions. Keeping up with these expansions is confusing and requires a lot of work. For example, if you want to use higher-order programming in TeX, you have to make sure that the function that is passed is not expanded immediately.

Next to this, we also have to watch out for name clashes, because in TeX, and LATEX, most of the macro definitions end up in the root namespace. The newcommand we saw in Subsection 1.1.2, requires that the command has not been defined earlier, if it was, an error is thrown. renewcommand requires that the command was defined earlier, it overwrites the previous command and it crashes if the command was not yet defined. Lastly, providecommand acts as newcommand if the command was not yet defined and does nothing if it was already defined, leaving the old definition intact. For example, we can not define a command called name using newcommand as this command already exists, we have to use renewcommand to this. This should be used with care however, as every command that uses name now uses the newly defined name.

Lastly, LATEX does not use a static type system while this could provide us with information needed for refactoring and auto-completion based on the context.

#### 1.2.3 Markdown

Markdown is simple but pays a hefty price for it. All of the syntax elements have been hard-coded into the language meaning there is no room for customisability. The only way to customise Markdown is by embedding inline HTML. This allows to create a solid structured model but it still does not allow for very much customisability. There is no programming model that can provide the customisability.

As an example, we use HTML to create a table, as these are not a part of default Markdown (though there are variations that have syntax for them, like GitHub Flavored Markdown [7]).

```
Below you can find an HTML table
1
2
  3
      Element 1
5
         6
7
         \langle t.d \rangle
             Element 2
          </t.d>
9
       10
11
```

#### Below you can find an HTML table

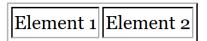


Figure 1.14: The table document rendered as HTML by the official Markdown conversion tool.

Adding HTML to Markdown decreases readability and cleanliness of your document significantly, which goes against the goal of Markdown. Using it also introduces the need for a certain technical proficiency to read and edit the source document. Finally, HTML reintroduces the possibility of syntax errors and thus slows down the document creation.

#### 1.2.4 Pandoc

Pandoc allows you to execute exterior programs through the PanPipe filter, but the only result is the stdout and stderr of the program. These do not allow you to change the output afterwards because you have no more structure. The structural elements such as sections and paragraphs have disappeared. For example, you create a binary tree using a 3rd party application. Then you want to insert a new value into this tree, which requires it to be rebalanced. However this is impossible as you lost all of the structure of the tree, it is only represented as a string. You thus have to redraw every instance of the tree instead of creating it once, programmatically transforming it and redisplaying the transformed tree.

To access the structure of the document, the AST can be used. Transforming an AST afterwards however, does not allow for precise control either. In Listing 1.3 we saw how we could replace all of the headers of a level higher than 2, but changing it for only a few would require us to find them. If we could have done this inline or if we could redefine the meaning of the # symbol, the problem would be avoided.

Finally, the variables used in Pandoc have some limitations. They can not be defined inline, only through the command line and they can only be used in templates. This means that

16

you can not easily embed variables and execute simple expressions such as x + y.

#### 1.3 Proposed solution: Neio

To counter the problems occurring in the state-of-the-art solutions whilst retaining their advantages, we created a new markup language called Neio (read as neo). The name Neio was chosen because neo means new and we are developing a new language. The spelling however, came from the Colemak keyboard layout [27], the letters on the home row, beneath the right hand, spell n-e-i-o.

Based on these solutions, the following goals were created for the Neio markup language:

- 1. It has to be user-friendly and easy to get started with the language;
- 2. It has to use a modern programming paradigm;
- 3. It has to be highly customisable.

The first goal is achieved by using a syntax like Markdown for the documents.

To achieve the second goal we note that a document has a strong structure consisting of a lot of elements like chapters, sections, paragraphs and so on. An object-oriented model, like the one in Java, is able to represent this structure well by using an object for each element and inheritance allows us to easily specialize an element, for example change the numbering of an enumeration. We also chose this model because it is well known and as seen in Subsection 1.1.2 and Subsection 1.1.4, working with macro expansions or filters is not ideal.

Because it offers a powerful programming model that we can translate to, we chose to translate to IATEX. It has also been used successively for decades and has proven that it can deliver very clean looking documents. Later on more back ends, such as HTML, could be implemented, but that is outside of scope of this thesis.

#### 1.3.1 Target group

Neio targets anyone that is currently using Markdown but wants to customise their document further. For example, using Neio you can create citations and references, which is not possible in Markdown.

By offering a simpler syntax than used in LATEX, we target LATEX users. Finally, we target developers by offering a well-known, modern programming model that is more conventional than the macro model in TeX.

## Chapter 2

# Design of the Neio markup language

In Chapter 1 the state-of-the-art concerning document creation has been discussed and we concluded that improvements can be made. In this chapter the Neio language is presented. We explain how our decisions were reached and how they were affected by the state-of-the-art.

#### 2.1 Neio document

The typical files that a user writes are called Neio documents; they have to be as simple as possible. These files use the .no extension. Based on our state-of-the-art analysis, we chose a Markdown-like syntax for Neio documents.

To illustrate some of the basic concepts we present an example Neio document.

2.2 Code files 19

Listing 2.1: A simple Neio document.

```
1  // This is a Neio document
2  [Document]
3
4  /* Below we define a chapter
5  * and a paragraph.
6  */
7  # Chapter 1
8  This is the first paragraph.
```

## 1 Chapter 1

This is the first paragraph.

Figure 2.1: The rendered form of the document to the left.

Even if you are not familiar with Markdown, you can immediately tell what this document represents. It creates a document, a chapter (through the # symbol) and a paragraph, and it contains two comments.

A difference with Markdown, but a resemblance with IATEX, is that every Neio document starts out with document class. It tells us what kind of document we are building. This is done to improve customisability and to support multiple kinds of documents.

In Neio there are two sets of syntax called text- and code mode. Listing 2.1 is written in text mode, the code files in next section are written in code mode.

#### 2.2 Code files

Code files are very similar to class files in Java. They also use the .no extension. To introduce code files, we take a look at the Paragraph class that is used to represent the paragraph in Listing 2.1.

Listing 2.2: Paragraph.no

```
namespace neio.stdlib;

import neio.lang.*;

/**

Represents a paragraph

*/
class Paragraph extends Content;

// private
```

2.2 Code files

```
11 Text text;
12
13
   * Intialises a paragraph with some text
14
15
   * @param parText The initial text of the paragraph
16
17
  Paragraph(Text text) {
18
       this.text = text;
19
20
21
^{22}
   * Appends some text on a new line within the same paragraph
23
24
   * Oparam t The text to add after the newline
25
  * @return This paragraph with a newline and {@code t} added to it
26
27
  Paragraph appendLine(Text t) {
28
      return appendText(new Text("\n").appendText(t));
29
30
31
32
   * Appends Text to this Paragraph
33
34
   * @param t The Text to add
   * @return This paragraph with {@code t} added to it
36
37
  private Paragraph appendText(Text t) {
38
      if (text == null) {
39
           text = t;
40
41
       } else {
42
           text = text.appendText(t);
43
44
       return this;
  }
45
46
47
   * Creates a newline handler to handle newlines following a Paragraph
49
   * @return A new {@code ParNLHandler} with this Paragraph as parent
50
51
  ParNLHandler newline() {
52
53
      return text;
54
55
56
  * Returns the LaTeX representation of this Paragraph.
   * It is the empty string in case there is no Text in this paragraph.
58
59
  * @return The LaTeX representation of this Paragraph
60
  String toTex() {
62
      if (text != null) {
63
          return "\\par " + text.toTex();
64
       } else {
65
          return "";
66
       }
67
  }
68
```

Except for some variations on the syntax, such as using namespace instead of package, and

2.3 Call chain 21

the lack of access level modifiers (which can be specified), this is valid Java code. The absent access level modifiers are automatically set to private for members and public for methods. They do not have to be written explicitly as the default value is usually what the developer wants and it makes the code file just a little clearer and more concise.

To keep code files simple and to maximize reusability (as well as making the parsing somewhat easier), some functionality from the Java language was dropped. It is for example not possible to create anonymous classes as these are not reusable and do not provide any new functionality. It is also not possible to create multiple classes in a single file.

As we see in the next section, code files are used to build a Neio document.

#### 2.3 Call chain

A problem with Markdown is that the semantics for the syntax are hard-coded in the language, a # always represents a chapter for example. But if you are creating a slide show for example, you want # to create a new slide instead of a chapter. For this reason, text mode in Neio has no hard-coded semantics.

Instead, the semantics of the syntax are defined in code files. For example, the # in a Document is defined as follows:

Listing 2.3: Document.no

```
Creates a Chapter and adds it to this
2
3
     Oparam title The title to use for the new Chapter
4
     Oreturn The newly created Chapter
5
6
  Chapter #(Text title) {
7
      Chapter chapter = new Chapter(title, 1);
8
      addContent(chapter);
9
10
      return chapter;
11
  }
12
```

Now we can see that a Neio document is actually a sequence of method calls. This is referred to as the call chain. An example of a document and its call chain is shown below.

2.3 Call chain

Note that the call chain is not code that the user has to write, its just a representation of a Neio document.

The user is thus actually creating an object model. The object model for the previous document is shown below. The calls that create every object are shown on the edges.

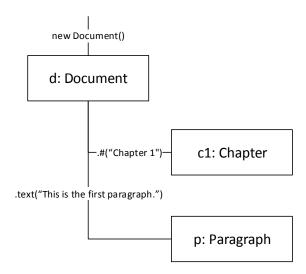


Figure 2.2: The object model of the previous document.

This object model is separated as much as possible from the visualisation, which happens later on. We say "as much as possible" because it is not completely separated. For example, sometimes it is specified that two elements should be placed next to each other in the object model.

Every element that is visualised in the final document is a Content. Content is the base class of Neio, it is like Object in Java.

To customise the semantics of the syntax in text mode, the Java syntax in code files has been modified to allow for special method identifiers. These are discussed in the next section.

23

## 2.4 Method modifications

To write the Neio documents we have seen so far, the regular Java methods used in code files had to be modified slightly. We also had to reserve two method names that are not reserved in Java. The modifications and the reasons for the modifications are given below.

## 2.4.1 Symbol methods

In Listing 2.3 we saw that the code file, used a hash symbol as a method name. These methods are called symbol methods. All the symbols that can be used as method names are #, -, \*, -, \$, |, =, ^, '.

These methods have either no arguments or a Text argument. This is done to keep the method calls invisible for the user as more or different arguments would have to passed explicitly. The Text that is passed to the method is the text following the call in text mode. An example of both, and the call chain, is shown below.

```
1 [Document] 1 new Document() 2 # Chapter 1 2 .#("Chapter 1") 3 # .#()
```

A symbol method can only be used at the beginning of a line or after another symbol method. The latter is used to chain methods together. We discuss applications of this in Chapter 3.

Next to symbols, there are two method names that have been reserved. The first of them is text and is explained in the next section.

## 2.4.2 Text

Whenever text in a Neio document is encountered on its own, for example a paragraph but not the text behind a # symbol, the text method is called with the text as an argument. text is a separate method to allow continuation of the call chain and because not every block of text is necessarily a paragraph. It can for example be a quote.

We use Text instead of just a Java String because Text can be marked up. How this can be done is explained in the next section.

## 2.4.3 Surround methods

Like in Markdown, we want to markup text by placing it between a pair of characters. For example, putting text in bold by surrounding it with a pair of star characters. With the features that we have seen so far this is not yet possible.

To achieve this surround methods were developed. A surround method is a method that is annotated with the **surround** modifier. When text in text mode is surrounded with a pair of symbols, then a surround method is called. A few examples of such methods are given below.

Listing 2.4: Text.no

```
1
  * Creates a new {@code BoldText} and appends it to this
2
3
  * Oparam t The text to make bold
4
  * Creturn The newly created BoldText
  public surround Text *(Text t) {
7
       return appendText(new BoldText(t));
8
9
10
11
  * Creates a new {@code ItalicText} and appends it to this
12
13
  * @param t The text to make italic
14
  * Oreturn The newly created ItalicText
15
16
17
  public surround Text _(Text t) {
       return appendText(new ItalicText(t));
18
19
20
21
  * Creates a new {@code MonospaceText} and appends it to this
22
23
  * Oparam t The text to show in a monospace font
  * Oreturn The newly created MonospaceText
25
26
  public surround Text '(Text t) {
27
28
       return appendText(new MonospaceText(t));
29
```

The text This is \*bold\* text is thus transformed into the following call chain text("This is ").text(\*(new Text("bold"))).text(" text"). The text call, wrapping the \* call,

2.4 Method modifications

25

is necessary because the \* method only creates a BoldText. It does not append it to the current text. Throughout the text the surround method ' is used a lot, it shows the text in a monospaced font.

As said in Subsection 2.4.1, the argument of symbol methods is a Text, this means we can use surround methods in a symbol method. A more extensive example of surround methods is shown below.

Listing 2.5: surroundEx.input

## 1 Chapter 1

This Chapter is written in italic and monospace font.

## 2 Chapter 2

This **Chapter** is written in italic.

Figure 2.3: The rendered version of the document to the left.

To not confuse surround methods with the symbol methods, the first and last characters inside a surround call can not be a space. The text after a symbol method on the other hand, is always separated from the symbol by at least one space.

## 2.4.4 Nested methods

Documents often have recursive elements such as sections or enumerations. With what we have seen thus far, we would have to create a new method for every level of recursion. For example, to create sections, subsections and subsubsections, we have to create the following methods: #, ##, ###.

This is of course very cumbersome and even impossible if there is no limit on the recursion. The behaviour of such recursive elements is also very similar. Usually, only one property, such as the numbering and indentation in an enumeration, is affected by the level of recursion. This is why nested methods were developed.

A method can be annotated with the nested modifier. Such a method implicitly takes an extra argument, an Integer that reflects the depth of this recursive method. The depth of this recursive method is the number of times the symbol has been used. If we have a look at the Chapter class, we see that it defines a nested method #.

Listing 2.6: Chapter.no

```
1
   * Creates a new Chapter and adds it to this.
   * If the level of the new Chapter is lower or equal
3
     to the level of this, the new Chapter is added to
   * the nearest TextContainer above this.
    * This is a nested call thus only ##+ will match.
7
     @param title
                    The title of the new Chapter
8
   * @param level The level of nesting of the new Chapter
9
10
  nested Chapter #(Text title, Integer level) {
11
       if (level <= this.level) {</pre>
12
           Chapter c = nearestAncestor(Chapter.class);
13
           if (c != null) {
14
               return c.hash(title, level);
15
             else {
16
17
               return nearestAncestor(Document.class).hash(title);
18
19
20
       Chapter chapter = new Chapter(title, level);
21
       addContent(chapter);
22
       return chapter;
23
24
  }
```

If we now call chapter.##("Chapter 1.1") for example, it translates to the following call: chapter.#("Chapter 1.1", 2).

A nested method only matches with at least two symbols, the single symbol has to be defined separately. The first reason for this is because the first level often requires some initialisation. A second reason is given once context types have been discussed. An example of such initialisation is shown below.

Listing 2.7: TextContainer.no

```
/**

2 * Creates a new Itemize and adds an ItemizeItem to it

3 * The Itemize is added to this.

4 *

5 * @param text The text to use for the ItemizeItem

6 * @return The created ItemizeItem

7 */
```

2.4 Method modifications

27

```
8    Itemize *(Text text) {
9         Itemize itemize = new Itemize();
10         ItemizeItem item = new ItemizeItem(text, itemize, 1);
11         itemize.*(item);
12         addContent(itemize);
13         return itemize;
15    }
```

When a \* call is encountered for the first time, an Itemize has to be created The following \* calls then simply add an ItemizeItem to it.

The last method name that was reserved is newline. It is explained in the next section.

## 2.4.5 Newlines

Users often use varying amount of newlines to express different structures. For example, the following two examples do not have the same meaning.

Listing 2.8: Two lists

Listing 2.9: One list

In Listing 2.8 there are two lists while in Listing 2.9 there is only one list. The same is done to separate paragraphs.

Because of this, and to include as much information as possible in the call chain, even the new-line character is a method. The newline character is defined by the following W3C Extended Backus-Naur Form (EBNF) [36]:

```
newline ::= "\r?\n"
```

An example of a document and its call chain including the newline method is shown below.

```
new Document()
                                                      .newline()
                                               2
  [Document]
                                                      .text("The first line")
  The first line
2
                                                      .newline()
                                               4
  of the first paragraph.
                                                      .text("of the first paragraph.")
3
                                               5
                                                      .newline()
                                               6
  The first line
                                               7
                                                      .newline()
  of the second paragraph.
                                                      .text("The first line")
                                               9
                                                      .newline()
                                                      .text("of the second paragraph.")
                                              10
```

This example shows that we never explicitly create a paragraph, this is done depending on the context. We explain the example step-by-step to understand how the document is build.

The first newline() is defined in TextContainer, the super class of Document, and creates an intermediate object that is an instance of NLHandler. The code for this method is shown below.

Listing 2.10: TextContainer.no

```
/**
    * Handles newlines

* * @return Returns a new NLHandler

*/
NLHandler newline() {
    return new NLHandler(this);
}
```

NLHandler defines the text method (shown below), which creates a paragraph and adds it to the parent, in this case the Document.

Listing 2.11: NLHandler.no

```
Creates a Paragraph and adds it to the parent.
2
3
     Oparam text The text for the Paragraph
4
     @return
                      The newly created Paragraph
5
6
  Paragraph text(Text text) {
7
8
       new Paragraph(text) par;
       parent().addContent(par);
9
10
       return par;
  }
11
12
13
   * Returns this
14
15
     @return this
```

```
17 */
18 NLHandler newline() {
19 return this;
20 }
```

The second newline() is defined in Paragraph. It creates a new ParNLHandler which also defines a text method. This text method (shown below) however does not create a new Paragraph but appends to the existing one.

Listing 2.12: ParNLHandler.no

```
/**
    * Appends some Text to an existing Paragraph

*    * Oparam text The Text to add

*    * Oreturn The existing Paragraph with {Ocode text} appended to it

*/
Paragraph text(Text text) {
    return parent().appendLine(text);
}
```

Next there are two newline() calls. As we saw the first one creates a new ParNLHandler, but ParNLHandler also defines a newline method (given below). This method returns the NLHandler defined by the parent (in this case Document) of the existing Paragraph.

Listing 2.13: ParNLHandler.no

The next text() is thus called on an instance of NLHandler and as said before, creates a new Paragraph. The following newline() and text() append some Text to the new Paragraph using a ParNLHandler, as seen before.

By introducing these special method identifiers we actually extended fluent interfaces [6]. Because of this, the user is actually programming without noticing it.

In Chapter 3 we see that we can for example create tables as the one below.

Listing 2.14: A table created using a DSL.

Ι	Student club	I	Rounds	1	Seconds/Round	I	Dist (km)	1	Speed km/h	I
1	HILOK		1030	·	42		298,70	- <b>-</b> -	24,89	 I
1	VTK	1	1028	- 1	42	-	298.12	-	24.84	-1
I	VLK		841		51		243.89		20.32	- 1
I	Wetenschappen and VLAK		819		53		237,51		19.79	-
	VGK	1	810		53		234.90		19.58	- 1
I	Hermes and LILA		793		54		229.97		19.16	
1	HK		771	- 1	56		223.59	-	18.63	
	VRG	1	764		57		221.56		18.46	
1	VEK		757	- 1	57		219.53	-	18.29	-
	VPPK	1	689		63		199.81		16.65	
	SK		647	- 1	67		187.63	-	15.64	-
1	Zeus WPI	1	567	-	76	-	164.43	-	13.70	-
	VBK	1	344		126		99.76		8.31	

Student club	Rounds	Seconds/Round	Dist (km)	Speed km/h
HILOK	1030	42	298,70	24,89
VTK	1028	42	298.12	24.84
VLK	841	51	243.89	20.32
Wetenschappen and VLAK	819	53	237,51	19.79
VGK	810	53	234.90	19.58
Hermes and LILA	793	54	229.97	19.16
HK	771	56	223.59	18.63
VRG	764	57	221.56	18.46
VEK	757	57	219.53	18.29
VPPK	689	63	199.81	16.65
SK	647	67	187.63	15.64
Zeus WPI	567	76	164.43	13.70
VBK	344	126	99.76	8.31

This table is not an image that was included, but is actually created as shown in Listing 2.14. We do not need to create a special reader or to add new syntax to implement this table. Instead we create a Domain Specific Language (DSL) to build tables using the features we have shown so far.

To avoid clutter, the newline method is not shown in the further examples unless it is needed

to understand the example.

## 2.5 Context types

Every method in the call chain returns an object on which the rest of call chain is called. However, we do not want to define the same symbol methods for all of these objects, for example defining # in a Document, a Chapter, a Paragraph and so on. We also do not want to specify on what object every method in the call chain is called. For this reason context types were developed.

A method call in text mode is not always called on the previously returned object. It can be called also be called on an another object which was defined earlier. Context types use the object model, that has been built so far, to find out which object roots the next method call.

We illustrate this using an example.

As we saw before, Document creates a NLHandler through the newline method. However, NLHandler does not have a # method. # in this example is called on Document instead and creates a Chapter.

Chapter also creates a NLHandler when newline is called on it and as we saw NLHandler defines text, which creates a Paragraph.

Paragraph creates a ParNLHandler when newline is called on it, but ParNLHandler has no # method. The last # is again called on Document. The call chain can be represented as a tree. This is shown below.

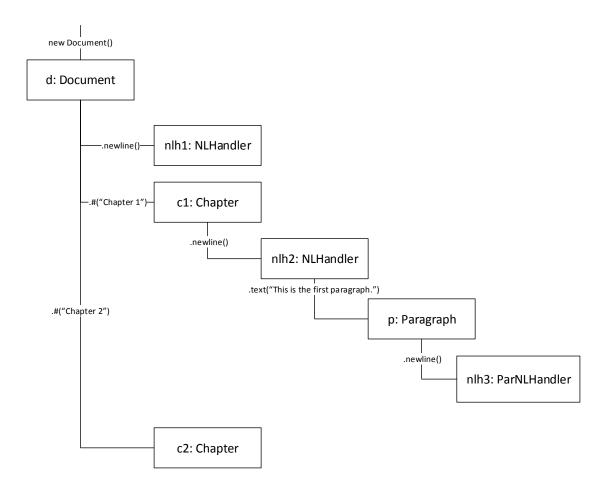


Figure 2.4: The tree representation of the call chain of the previous example.

A context type is created every time a new method call (in the call chain) returns. It has a context and an actual type. The latter is the type of the object that was returned by the new method call. The former is the context type in which the type that defines the method of the method call was found.

If the previous context type defines the method of your method call, but the actual type does not, then we search for the method in the context.

To understand how context types find the object on which a method is called, we go over the previous example step-by-step .

When we encounter new Document(), the first context type, called ct0, is created.

There are no context types yet, so there is no context and the actual type is **Document**. The context type is shown below.

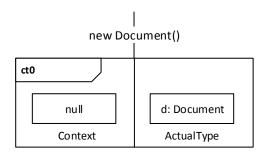


Figure 2.5: The first context type: ct0.

When we now encounter newline() we have to do a context lookup. This lookup recursively searches through previous context types for the object that roots the method call.

The context lookup starts by checking if the actual type of the previous context type, ct0, defines a method newline. As it does, we do not have to check the context of ct0.

The context of the new context type, ct1, is ct0 and because newline() creates a NLHandler the actual type is NLHandler. The tree representation of the call chain, with context types, now looks as follows.

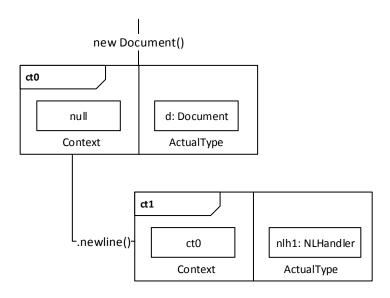


Figure 2.6: The context types ct0 and ct1.

The next call is #("Chapter 1") and the previous context type is ct1. To know on what to call # we first check if the actual type of ct1 defines a # method. It does not, thus we check if the context of ct1 does so instead.

The context of ct1 is ct0, thus we first check if the actual type of ct0 defines #. The actual type is Document and this defines #. The new context type, ct2, is thus created with ct0 as context. Because the # in Document returns a Chapter, the actual type is Chapter. The tree representation is shown below.

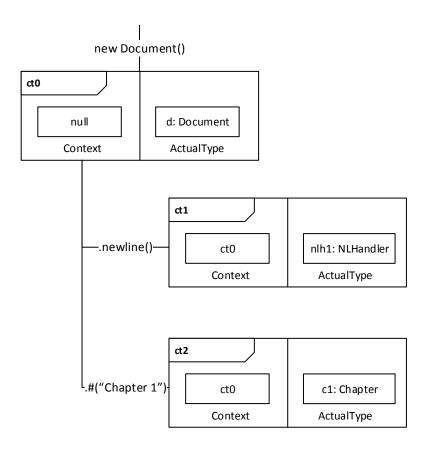


Figure 2.7: The context types ct0 through ct2.

The following call is newline(), we check the actual type of ct2, Chapter, and we know that it defines newline. We thus create ct3 with ct2 as context and NLHandler as actual type. ct4 and ct5 are created with respectively ct3 and ct4 as context and respectively Paragraph and ParNLHandler as actual type. The tree representation of ct0 through ct5 is shown below.

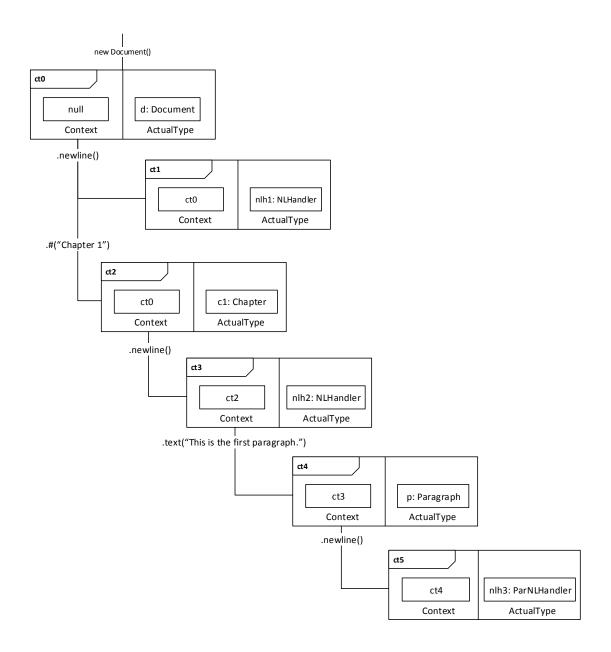


Figure 2.8: The context types ct0 through ct5.

Finally, #("Chapter 2") is called. For this call we recursively search back up to ct0 as it is the first context type in which # is defined. (Par)NLHandler, Paragraph and Chapter have no definition of #, only Document has such a definition. The last context type is thus built with ct0 as context and Chapter as a actual type. The final tree representation is shown below.

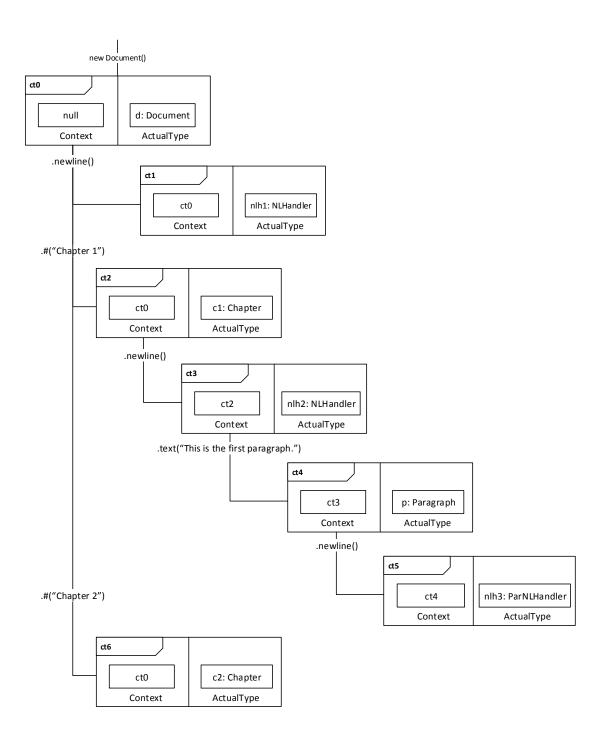


Figure 2.9: The context types ct0 through ct6.

It is important to note that we do not carry the entire context along for the entire document. In the last step of the previous example, a context reset happened as part of the lookup

procedure. After a context reset a part of the object model becomes inaccessible for future method calls in the call chain. This is needed to correctly create the object model.

To show why context resets are important we give another example:

Listing 2.15: An example for context resets.

```
[Document]
2 * Item 1
3 * Item 2
4
5 Paragraph 1
* Item 3
```

```
new Document()
2    .*("Item 1")
3    .*("Item 2")
4    .newline()
5    .text("Paragraph 1")
6    .*("Item 3")
```

The example has the following object model. The methods that cause a context reset have been coloured red.

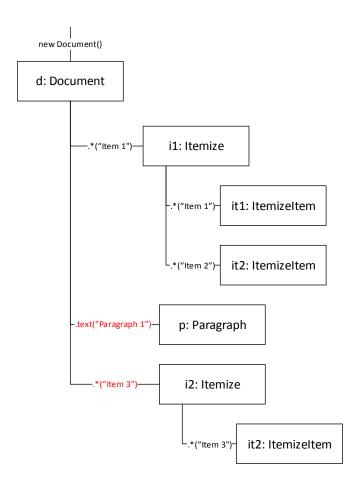


Figure 2.10: The object model of the previous example.

The \* is defined in Document and creates Itemize i1. The \* defined in Itemize creates ItemizeItems and adds them to the Itemize. Because of context resets, the last \* is called on Document. This document thus contains an itemize of two items, a paragraph and an itemize of one item. The correct context lookup for the last \* is visualised on the call chain below, the arrows depict one step in the context lookup.

Figure 2.11: Correct context lookup for \*.

If there were no context resets and the entire context was thus kept in every step, Item 3 would find the \* defined in Item 2 before it reached Document. This means that there would only be one list, that contains all three items, instead of two lists as was meant in the document. This wrong context lookup for the last \* is visualised on the call chain below.

```
new Document().*("Item 1").*("Item2").newline().text("Paragraph 1").*("Item 3")
```

Figure 2.12: Wrong context lookup for \*.

## Nested methods

Now that we know how context types work, we can explain why the first level of a nested method is separated from the rest. We explain why this is needed using the following example:

```
1 [Document]
2 # Chapter 1
3 ## Chapter 1.1
4 Chapter 2
1 new Document()
2 .#("Chapter 1")
3 .#("Chapter 1.1", 2)
4 .#("Chapter 2")
```

We want Chapter 1 and Chapter 2 to be children of Document and Chapter 1.1 should be a child of Chapter 1. The context type representation of the document is shown below.

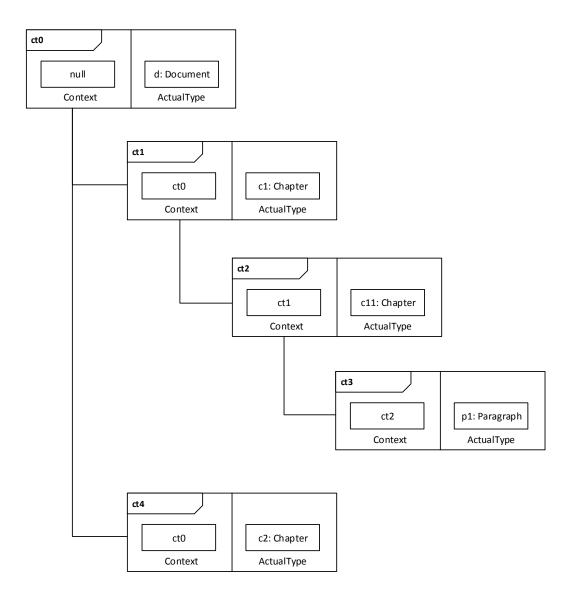


Figure 2.13: The correct context type representation.

If we did not separate the first level, the aforementioned structure is still built correctly, **but** the ContextType representation is different. This is because we did not reset to the same point, the context reset jumped up to Chapter 1 instead of Document. The context type representation of this document is shown below.

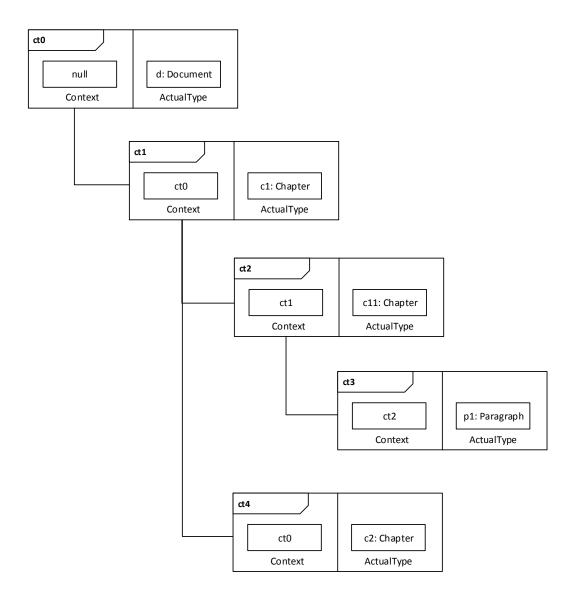


Figure 2.14: The wrong context type representation.

## 2.6 Code blocks

We have seen that we can use code in code files and customise our document this way. But this does not allow us to do everything we want to. We want to define variables and execute short expressions such as x + y We do not want to create a new document class every time

we want to do something out of the ordinary.

We want to change properties of content in a Neio document without touching code files. It also has to be possible to insert content in such a document without using a symbol method. For example, we can create a class for pie charts but as there is no symbol method that creates it in the existing document classes, we can not insert it with the features discussed thus far.

For this reason, it is possible to add code blocks in a Neio document. The three different kinds of code blocks are explained below.

## 2.6.1 Non-scoped code

The first kind of code blocks is the non-scoped code block. We show how it is used with an example.

In the example above a code block is opened and a new Chapter is created. A non-scoped code block is defined by the following EBNF:

The statements in such a block are injected directly into the document, without introducing a scope. This means that we can use variables defined in a non-scoped block, later on in the document. However, this means that we should be careful when choosing names for our variables as they are added to the global namespace of the Neio document.

We have now created a Chapter but it is not part of the document yet. We could add it manually by calling addContent(), but there is an easier way as this is a pattern that occurs a lot.

If the last statement in a non-scoped code block returns an object, this object is appended to the current document. We can thus also use a return statement as last statement to add an object to the document. These two possibilities are shown below.

Listing 2.16: Automatic return from code block.

Listing 2.17: Manual return from code block.

```
[Document]
                                                    Chapter 1
                                                2
  [Document]
                                                3
                                                   {
                                                       Chapter chapter2 = new Chapter("
  # Chapter 1
2
                                                4
3
  {
                                                           Chapter 2");
4
      new Chapter("Chapter 2")
                                                5
                                                       return chapter2;
  }
                                                   }
5
                                                6
```

The object returned in the last statement is added to the document and it becomes part of the context. It is added to the document by calling appendContent on this. To further explain code blocks we explain this in the next section.

#### 2.6.2 This

Like every object-oriented language, Neio has an expression to refer to the current object. Because the document is actually a chain of method calls, it is set to the last object returned before a code block. When an object is returned from a code block and added to the document, this object becomes the new this.

We need this as we do not know how to refer to the objects that were created in text mode. The example in Listing 2.16 can be seen as the following call chain.

```
new Document()
    .#("Chapter 1");
Chapter chapter 2 = new Chapter("Chapter 2")
this.appendContent(chapter 2);
```

A separate statement is created for the call chain before a code block. The object returned from the code block is placed in a variable and passed to the appendContent method call. When Listing 2.16 is actually translated, this is be filled in and it produces the following code:

```
Document $var0 = new Document();
Chapter $var1 = $var0.#("Chapter 1");
```

```
Chapter $var2 = new Chapter("Chapter 2");

$var1.appendContent($var2);
```

this is filled in with the last object that was returned in the call chain before the code block. To do this the compiler has knowledge of the appendContent method.

We show another example to show how context types and this work together. For this example we use the following document:

In this example we create a chapter as we do in text mode, by using the # method. Chapter has no # method, but has Document does. The method already adds the created Chapter to the Document. Thus we assign a value to the result of the # method to prevent it from being automatically appended like it was in the previous examples.

Because of context types this is be filled in by the variable that holds Document instead of the last object that was returned in text mode. The final translation is given below.

```
Document $var0 = new Document();
Chapter $var1 = $var0.#("Chapter 1");
Chapter chapter2 = $var0.#("Chapter 2");
```

## 2.6.3 Scoped code

Sometimes all we want to do is execute some arbitrary code without corrupting the namespace.

For this, we use a scoped code block. Its use is illustrated below.

```
[Document]
  # Chapter 1
2
  {{
3
  List < String > 1 = new ArrayList < String > ();
  1.add("upquote");
  1.add("pdfpages");
  1.add("url");
       (int i = 0; i < 1.size(); i = i + 1) {
       addPackage(1.get(i));
9
10
  }
  }}
11
```

In this example a Document is created, but to suit our needs, it needs a few more packages that are not included by default in Document. We also do not want to think of very descriptive identifiers for our variables as it is just a simple operation, the name 1 should suffice.

A scoped block does the same as a non-scoped code block but it injects all of the code in a new scope. The last statement is not automatically appended to the document, to be certain that nothing leaves the scope. A scoped block can thus be used as a safe way to execute code without adding anything to the document, changing the current this or polluting the global namespace.

The call chain of this example is given below.

```
new Document()
       .#("Chapter 1");
2
  {
3
      List<String> 1 = new ArrayList<String>();
4
5
      1.add("upquote");
      1.add("pdfpages");
6
      1.add("pdfpages");
       for (int i = 0; i < 1.size(); i = i + 1) {
8
           this.addPackage(1.get(i));
9
10
11
```

Scoped blocks have been used frequently while creating this document. One of its prominent uses has been to add images. This is because TextContainer defines an image method that already adds the image to the document. As we do not want to add the object twice, we use scoped code blocks to call this method. The image method and and example of how images are typically included is shown below.

## 2.6.4 Inline code

The the last type of code blocks are inline code blocks. As the name implies, this is code that is meant to be used inline. As such it can only contain one statement and the opening and closing bracket have to be on the same line.

Inline code blocks can be used anywhere that text can be used. In the example below we show how inline code blocks can be used in combination with a non-scoped code blocks to create a template for a letter.

2.7 Text in code mode 47

Listing 2.18: template.no

```
[Document]
1
2
                            "Thomas Vanhaskel
        Text addressee =
3
        Text help = "my math class";
4
        Text helpSubject = "math test";
5
        Text closings = "Kind regards,";
Text name = "Titouan Vervack";
6
7
   }
8
   Dear {addressee},
10
11
   Thank you for helping me out with {help
12
13
   I was able to do great at the \{
14
       helpSubject} thanks to you.
15
   {closings}
16
17
   {name}
18
```

Dear Thomas Vanhaskel,

Thank you for helping me out with my math class.

I was able to do great at the math test thanks to you.

Kind regards,

Titouan Vervack

Figure 2.15: The rendered form of the example to the left.

An inline code block returns a text and is appended to the rest of the document by calling the text method on this.

## 2.7 Text in code mode

In the previous sections we created objects such as Chapters as follows new Chapter ("Chapter 1"). However, "Chapter 1" is not a Java String it is a Text. Text can thus be written in code mode by enclosing text with a pair of doubles quotes (").

A Java String is can still needed and can therefore be created using a pair of triple, single-quotes ('''). One of the most prominent reason for using String instead of Text is when you have to pass a path for example to pass code to a code listings.

The inspiration for using three quotes came from the multi-line String literal in Python. The reason we use three single quotes, instead of only, is because a pair of single quotes still represents a Java Character. We do not use two quotes because that that would be very confusing when used in combination with double quotes in a non-monospace font.

As we saw in Subsection 2.6.4, you can use code blocks in text mode. Since we can also use text mode in code mode, this means we can endlessly nest text and code mode.

An example demonstrating the strength of this nesting is shown below.

Listing 2.19: dice.no

```
[Document]
1
2
  We roll the dice!
3
4
5
  {
       Random rng = new Random();
6
       Integer limit = 12;
       Integer random = rng.nextInt() %
8
           limit;
       if(random > (limit / 2)) {
9
           "Ouch,
10
                  we lost {random -
                                       (limit
                / 2)}!"
       } else {
11
           if (random == (limit / 2)) {
12
                "Oof, we broke even!"
13
           } else {
14
                "Great! We won {(limit / 2)
15
                     - random }!"
           }
16
       }
17
  }
18
```

## We roll the dice!

# Great! We won 8€!

Figure 2.16: The rendered form of the example to the left.

In this example we play a game where we roll two dice. In case we roll more than half of the total possible amount (12 in this case), we win.

In every case of the if statement, a piece of Text is found. As we saw in Subsection 2.4.2, this is translated to a text call. Since we placed and empty newline before the code block, this text is called on a NLHandler and a new paragraph containing the results of the game is created.

## 2.8 Navigating through the lexical structure

We have discussed all the features of Neio and we know how the object models are build. The lexical structure represented by the object model can now be navigated using a few methods in Content.

For example, we can use the nearestAncestor and directDescendants methods to count the number of top-level chapters in a document. An example of this is shown below as well as the code for the two methods.

```
[Document]

# Chapter 1

## Chapter 2

# Chapter 2

This document contains {nearestAncestor (Document.class).directDescendants(Chapter.class).size()} top-level chapters.
```

## 1 Chapter 1

## 1.1 Chapter 2

## 2 Chapter 2

This document contains 2 top-level chapters.

Figure 2.17: The rendered version of the example to the left.

Listing 2.20: Content.no

```
return list;
1
  }
2
3
4
   * Recursively checks for a parent of type {@code c}
5
6
   * Oparam c Type of the demanded Content
7
   * Oreturn The first lexical parent of type {Ocode c}
8
9
  <T> T nearestAncestor(Class<T> c) {
10
11
       Content p = parent();
       while ((p != null) && (!c.isInstance(p))) {
12
           p = p.parent();
13
14
15
       return list;
  }
16
17
18
   * Returns all direct descendants (the ones right beneath this) of
19
   * class {@code k}
20
21
   * @param k The class of the descendants we are looking for
22
     Oreturn A list of direct descendants of class {Ocode k}
23
24
  <T extends Content> List<T> directDescendants(Class<T> k) {
25
26
       List<T> descendants = new ArrayList<T>();
       for (int i = 0; i < contentSize(); i = i + 1) {
27
           Content c = content(i);
28
           if (k.isInstance(c)) {
29
               descendants.add((T) c);
30
           }
31
       }
32
```

2.9 Considerations 50

## 2.9 Considerations

Whilst constructing Neio, a few other things were considered but not implemented. They are discussed below.

## 2.9.1 Static typing

Neio has been designed to use static typing. The reason for this is, that while we are writing a Neio document in Code mode, or we are writing a Neio class, we can be notified of possible type mismatches. This decreases the time spent debugging and spent waiting on the compiler as you weed out one type mismatch after another.

It also allows you to create a safer program, as type errors are caught at compile time. On the other hand when using dynamic typing, type errors only show up at runtime. Using auto-completion in an IDE, we also do not have that much more typing to do than when we would be using dynamic typing.

The types have been incorporated into the Neio compiler, but the necessary checks to create comprehensive error messages are not yet in place. In the current compiler, a LookupException is thrown whenever there is a type mismatch, undefined variable,... as the method or variable that we are looking for can not be found.

## 2.9.2 Security

Another issue that we have considered is security. This is an issue because it is possible to directly execute Java code from the Neio document. However, security is also an issue in LATEX, as has been shown in several articles [3] [4].

Since Neio is more readable than LaTeX, malicious code can be found more easily than in LaTeX.

Adding a layer of security on top of Neio would require a lot of effort and research and is not in scope of this thesis. Thus we conclude by saying that we known there is a security risk,

2.9 Considerations 51

but dismissing it is seen as future work.

# Chapter 3

# Supported document types and libraries

This chapter goes into more detail about what document types can be handled by Neio and how to specifically build these kind of documents. We also show the flexibility of Neio by discussing some libraries that have been developed.

## 3.1 Document classes

Before we discuss the developed document classes it is important to note that there is still a lot more work has to be done on them. If we just have a look at the user guides of Memoir [18] or KOMA-Script [12], two popular document classes in LATEX, we see that they consist of 609 and 419 pages respectively. This is not something we could hope to reproduce in the scope of this thesis.

## 3.1.1 Document class definition

The document class to be used in a Neio document is defined by placing [DocumentClass] at the top of the file. With this line you are specifying what the first object in a Neio document

has to be. This object is the root of the call chain.

However, a document class definition is actually an expression that returns an object of this type. For example, when you write [Document], what you get as root of the call chain might be a Document. It might also be an instance of the Thesis document class, discussed in the next section, as Thesis is a subclass of Document.

The reason that this could be also be a subclass is because documents can also be included in another document. To include a document we can pass the main document to the included document. The included document then uses the main document as root of its call chain.

Because of this it is good practice to choose the most generic document class that is available for your document.

#### 3.1.2 Document

The simplest of documents can be created using the Document document class. It provides the syntax to create anything that Markdown can create, and on top of that it allows you to create tables, LATEX math, UML,.....

Like every Neio document, it can also use the code blocks defined in Section 2.6 to customise your document. For example, it is possible to place two Contents next to each other by using the leftOf and rightOf methods that are defined in the Content class. The code of these methods is shown below.

Listing 3.1: Content.no

```
1
   * Sets {@code c} right of this Content and sets the parent of
2
3
     {@code c} if we have a parent or sets this parent if {@code c}
4
   * has a parent. One of the Contents should have a parent!
5
   * Oparam c The Content to our right
6
     Oreturn The Content to our right
  <T extends Content> T leftOf(T c) {
9
      if (c == null) {
10
          return c;
11
12
      Minipage mp = prepareParent(this, c);
13
      Integer siblings = ((Minipage) (mp.content(0))).siblings();
14
```

```
Minipage mp2 = new Minipage(siblings);
15
       mp2.addContent(c);
16
       mp.addContent(mp2);
17
18
19
       return c;
  }
20
21
22
   * Sets {@code c} right of this Content and sets the parent of
23
24
   * {@code c} if we have a parent or sets this parent if {@code c}
   * has a parent. One of the Contents should have a parent!
25
26
   * Oparam c The Content to our left
27
   * Oreturn The Content to our left
28
29
  <T extends Content> T rightOf(T c) {
30
       c.leftOf(this);
31
       return c;
32
  }
33
34
35
36
   * Prepares the parent of {@code left} or {@code right} to have
37
   * horizontally aligned children (by using Minipage).
38
   st Oparam left. The left element whose parent has to be prepared
40
   * @param right The right element whose parent has to be prepared
41
   * Oreturn The root Minipage
42
43
  private Minipage prepareParent(Content left, Content right) {
44
       Content child = null;
45
       if (left.parent() == null) {
46
           child = right;
47
48
       } else {
           child = left;
49
50
51
       if (Minipage.class.isInstance(child.parent())) {
52
           List < Minipage > descendants = child.parent().directDescendants(Minipage
53
               .class);
           for (int i = 0; i < descendants.size(); i = i + 1) {</pre>
54
               Minipage m = descendants.get(i);
55
               m.incrementSiblings();
56
57
           return (Minipage) (child.parent().parent());
58
59
60
       Minipage rootPage = new Minipage(0);
61
       child.replaceWith(rootPage);
62
63
       Minipage mpl = new Minipage(1);
64
       mpl.addContent(left);
65
       rootPage.addContent(mpl);
66
67
68
       return rootPage;
69
```

This code takes the parent of one of the two Contents, the base, and links the other Content to the base. The implementation creates a root LATEX Minipage [20] if none of these object

is already next to another object. Otherwise it just wraps the two Contents in a Minipage. The root Minipage takes up the total width, while the Minipages that wrap the content take up (1/children) of the width.

Another way of implementing this, that would probably have worked better, was through TikZ pictures. It would have been better because it allows you to easily set any two things next to or above/beneath each other. Minipages require a lot more manual work and they might also split at a page-end which is not what wanted.

The Document document class can thus be used to write simple documents, such as small reports. Another document class that is complexer and requires some configuration, is shown below.

## 3.1.3 Book

To write a book, such as this thesis, we need a more complex document class. To write this thesis we created a new document class that represents the LATEX template for a thesis.

In this case the template is implemented in the Thesis class given below. To reuse as much functionality as possible, it extends the Document class. It adds its own packages and required elements, such as a title page, and it creates its own IATEX preamble. The code for this is shown below, the preamble is not shown entirely as it is very long.

Listing 3.2: Thesis.no

```
Intialises the document class, adding the required packages and adding a
2
       ToC
   */
3
  Thesis() {
4
       addPackage("graphicx")
5
       .add("fancyhdr")
6
       .add("amsmath")
7
       .add("float")
8
       .add("listings")
       .add("tocloft")
10
       .add("color")
11
       .add("pdfpages")
12
       .add("hyperref");
13
14
       List<String> options = new ArrayList<String>();
15
       options.add("english");
16
```

```
addPackage(options, "babel");
17
18
       addClassMapping(Chapter.class, ThesisChapter.class);
19
       addContent(new Pdf("TitlePage"));
20
       addContent(new BlankPage());
21
      addContent(new Pdf("TitlePage"));
22
       addContent(new FrontMatter());
23
  }
^{24}
25
26
   * Builds the preamble for this document class
27
28
     Oreturn The preamble for this document class
29
30
  String preamble() {
31
       StringBuilder preamble = new StringBuilder("\\documentclass[11pt,a4paper,
32
          {\tt oneside,notitlepage]\{book}\n")
       .append("\\setlength\{\t opmargin\}\{2cm\}\n")
33
       .append("\\setlength{\\headheight}{14.49998pt}\n")
34
       .append("\\setlength\{\headsep\}\{1cm\}\n")
35
       36
       .append("\\setlength\{\\evensidemargin\}\{2.5cm\}\n")
37
       .append("\\setlength\{\textwidth\}\{16\,cm\}\n")
38
       .append("\\setlength\{\textheight\}{23.3cm\}\n")
39
       .append("\\setlength\{\footskip\}\{1.5cm\}\n")
       .append(packages().toTex())
41
42
       return preamble.toString();
43
```

In the above code we also tell the document class to not create Chapters, but instead create ThesisChapters. A ThesisChapter redefines a chapter such that a level one chapter is translated to a LATEX chapter instead of a LATEX section. This is done through the following line.

```
addClassMapping(Chapter.class, ThesisChapter.class);
```

The document class also provides a method that initialises the main matter, which contains a Table of Contents (ToC) using the following line. This is a method that has to be called in the Neio document when the main matter begins. It is a separate method to allow us to easily add content, such as an abstract, before the main matter.

## 3.1.4 Abstract

The abstract used in this thesis is written in a separate Neio document (overview.no). It uses the Abstract document class and as such we import it into this document as follows.

Listing 3.3: thesis.no

```
overview.createDocument(parent().addContent(new Abstract()));
```

The first part of the Neio document for the abstract is shown below.

Listing 3.4: overview.no

```
[Abstract]
  | The design and implementation of a userfriendly object-oriented markup
2
      language
  | Titouan VERVACK
  | Science in Computer Science Engineering
  * Prof. dr. ir. Marko van Dooren
  - Dr. ir. Benoit Desouter
  | Engineering and Architecture
  | Ghent University
  | Applied mathematics, computer science and statistics
10
  | Prof. dr. Willy Govaerts
11
  # Abstract
12
  A lot of solutions exist for document creation.
```

As shown in the code above, the parameters for the Abstract are passed through symbol methods. To do this, Abstract overrides the newline method such that it creates an AbstractNLHandler. The AbstractNLHandler defines the |, - and \* methods. It also overrides the newline method to return the control of newlines back to the enclosing Content. The code for the - and | methods is shown below.

Listing 3.5: AbstractNLHandler.no

```
1
   * Adds a counsellor to the Abstract
2
3
   * Oparam text The counsellor to add
   * Oreturn The Abstract that created this NLHandler
5
6
  Abstract -(Text text) {
7
       parent().counsellor(text.realText());
8
9
       return parent();
10
11
12
   * Fills in a new parameter depending on what has been filled in already
13
14
   * Oparam text The value of the parameter
15
   * Creturn The Abstract that created this NLHandler
16
^{17}
  Abstract | (Text text) {
18
       Abstract a = parent();
```

```
if (a.title() == null)
20
21
           a.title(text.realText());
         else if (a.author() == null)
22
           a.author(text.realText());
^{23}
        else if (a.program() == null)
24
           a.program(text.realText());
25
         else if (a.faculty() == null)
26
27
           a.faculty(text.realText());
        else if (a.university() == null) {
28
29
           a.university(text.realText());
         else if (a.department() == null)
30
           a.department(text.realText());
31
         else if (a.chairman() == null) {
32
           a.chairman(text.realText());
33
34
35
       return parent();
36
37
```

The implementation of | is not perfect and the way to handle such scenarios should be changed in the future.

We could also have set the parameters using a code block and calling setters defined in Abstract. However, this seemed like a more elegant way.

### 3.2 Libraries

To demonstrate what the strength and flexibility of Neio a few libraries were created or reimplemented. All of the currently available libraries in Neio are explained below.

Note that none of these libraries are complete and it would require a lot more effort to reach a state of completion. For example, if we have a look at the manual of a popular library such as TikZ[26], we see that more time is needed to write such a library than there is available for this thesis.

### 3.2.1 BibTeX

The library created for BibTeX is one of the easiest ones created using Neio. The implementation is actually a binding to LATEX and does not allow for a lot of customisation as we do not have an object model of the BibTeX file.

The first reason we chose to implement BibTeX this way, is to illustrate that creating a binding for a LATEX package is not difficult. The second reason is that we would have had to create an entire BibTeX parser to create a complete object model of the BibTeX files. As the BibTeX format is quite complicated we felt that our time was better spent elsewhere.

The important code for our implementation of BibTeX can be found below.

Listing 3.6: Bibtex.no

```
1
   * Cites something defined in this BibTeX
3
     Oparam cname The keyword of the citation
     @return A Citation object representing the citation of {@code cname}
5
6
  Citation cite(String cname) {
7
       return new Citation(cname);
8
9
  }
10
11
   * Creates a TeX representation of this BibTeX, without citations
12
13
     Oreturn The TeX representation of this
14
15
  String toTex() {
16
       StringBuilder result = new StringBuilder("\n\\bibliographystyle{plain}\n")
17
       result.append("\\bibliography{").append(name).append("}\n");
18
19
       return result.toString();
20
21
```

The toTex() method creates the LATEX representation of a Content. This method is further discussed in Chapter 4.

A BibTeX file can now be added through the addBibtex method and we can cite an entry by using the cite method. These methods are both implemented in Document and are shown below.

Listing 3.7: Document.no

```
/**
2  * Sets a BibTeX file for this Document
3  *
4  * @param The name of the BibTeX file
5  */
void addBibtex(String name) {
  if (name != null) {
    this.bibtex = new Bibtex(name);
}
```

```
}
9
10
  }
11
12
   * A proxy method for {@code BibTeX}'s {@code cite}
13
   * Returns a Cite that is linked to {@code key}
14
15
   * Oparam key The key to access the citation
16
      Oreturn The Cite corresponding to {Ocode key}
17
18
  Citation cite(String key) {
19
       if ((bibtex != null) && (key != null)) {
20
           return bibtex.cite(key);
21
       } else {
22
           return new Citation("");
23
24
  }
25
```

The Citation class is a subclass of Text as we call it through an inline code block. This is shown below.

```
[Document]
This is how we cite something {cite('''key''')}.
```

key is a Java String as it is used to do a lookup on the BibTeX entries, it is thus not allowed to be marked up.

### 3.2.2 References

The way references have been implemented is quite straightforward. Any class that implements the Referable interface, such as Chapter or Image, can be referenced by calling the ref method defined in Referable. An example of referring to a chapter is shown below.

```
[Document]

# Chapter 1

(Chapter chap = (Chapter) parent())

In {chap.ref()} we explain references.
```

The example makes use of the parent method, defined in NLHandler, to put the newly defined Chapter in a method.

A NLHandler takes the object that created it as a parameter in its constructor so that it can refer to it later on. The newline method of a Chapter is defined in TextContainer (the super class of Chapter). The parent method in the example thus returns a TextContainer,

hence the cast to Chapter.

To reference an object, LATEX needs a unique label. We use the hashcode of an object for this as it is unique and easy to retrieve at any time. The ref method defined in Chapter and the Referable interface are shown below.

Listing 3.8: Chapter.no

```
Creates a reference to this Chapter
2
       and adds a correct prefix to it
3
      such as Chapter, Section or
       Subsection.
      Oreturn A Reference to this Chapter
5
6
  Reference ref() {
7
       String prefix = "";
       if (level < 2) \{
9
           prefix = "Chapter";
10
        else if (level == 2) {
11
            prefix = "Section";
12
13
        else {
           prefix = "Subsection";
14
15
16
       return new Reference(prefix, "" +
          hashCode());
18
```

Listing 3.9: Referable.no

```
namespace neio.lang;
1
3
   * Declares that a class can be
       referred to
5
  interface Referable;
6
7
   * References an object
9
10
     Oreturn A reference to this object
11
   */
  Reference ref();
13
```

As we need a variable to reference something, this means that the object we want to refer to has to be available when we want to refer to it. This means that we can not refer to things that are defined later on in the project, which LATEX can do.

IFTEX is able to do this because it compiles more than once and is able to gather the information needed to create a reference in one of the earlier runs. To do this in Neio, we would have to hold of creating the reference until the document is completed.

### 3.2.3 LATEX math and amsmath

INTEX is very good at typesetting mathematical formulas, as such we could not forget about it.

Two ways to make use of the LATEX math and the amsmath package have been created. The UML below shows the most important classes in the Neio math library.

```
Sqrt

Sqrt(arg: Content): Sqrt

Sqrt(root: Content, arg: Content): Sqrt

toTex(): String
```

```
Eq

Eq(): Eq

sqrt(root: Content, arg: Content): Eq

sqrt(arg: Content): Eq

(base: Content, pow: Content): Eq

nonu(): Eq

= (): Eq

v(t: Text): Eq

toTex(): String
```

```
InlineEq
InlineEq(): InlineEq
sqrt(root: Content, arg: Content): InlineEq
thisToTex(): String
```

An inline math formula in LaTeX is created by surrounding a formula with a pair of 's. In Neio we can create inline math using an inline code block. We can also wrap the code block with a \$ surround method, that does nothing, to make it look like inline math in LaTeX. However, this does not offer any new functionality and it is thus not done in this example.

The only method that has been implemented for this class is sqrt, so we show how we can create an inline square root.

The ieq method, defined in Document, creates an InlineEq. On this object we call the sqrt method. However, the sqrt method has Text arguments. To transform the Integer variables to Text, we enter text mode and open an inline code block that returns the variable. A solution to this problem is proposed in Chapter 5.3.

It is also possible to explicitly create an equation, in LATEX this is done using the equation environment. An example of how to create this in Neio is shown below.

$$\sqrt[2]{10}$$

$$2^{10} = 1024.0$$
(1)

Figure 3.1: The rendered form of the example to the left.

The code above creates two Equations. We disabled numbering on the first one through the nonu method and called sqrt on it.

In the second equation we use the ^ method to create an exponentiation and use the = method to create an equation. The v method creates a Value and calculates the actual value of what has been typeset before it. To be clear, everything except for the Math.pow() typesets something, but computes nothing. This can be clearly seen in Figure 3.1.

### 3.2.4 LATEX tables

Tables are one of the main things missing in Markdown and is one of the things people have the most problems with in LATEX. In this section we explain how the table shown in Subsection 2.4.5 was built.

To create tables we created a DSL that is readable in text mode and that allows you to easily edit values in the table later on. The latter is important because customising table elements inline, as is done in LATEX, makes a table hard to read.

The table is created as an ASCII table, with pipes, spaces, dashes and values for the elements of the table. The example shown in Subsection 2.4.5 is repeated below.

Listing 3.10: An ASCII table created in Neio.

1	Student club	I	Rounds	I	Seconds/Round	1	Dist (km)	1	Speed km/h	1
1	 HILOK	. <u>-</u> .	1030	· I	42	 I	298,70	 I	24.89	 I
i	VTK	İ	1028	i	42	Ĺ	298.12	i	24.84	İ
1	VLK	1	841	1	51	$\perp$	243.89	Ι	20.32	1
1	Wetenschappen and VLAK	1	819	-	53	-	237,51	-	19.79	-
1	VGK	1	810	-	53	-	234.90	-	19.58	-
$\perp$	Hermes and LILA	1	793		54	-	229.97		19.16	
	HK	1	771		56	-	223.59		18.63	
	VRG		764		57		221.56		18.46	-
1	VEK	1	757		57	-	219.53		18.29	-
1	VPPK		689		63		199.81		16.65	
1	SK	1	647	- [	67	$\perp$	187.63	-	15.64	-
1	Zeus WPI	1	567	-	76	-	164.43	-	13.70	-
1	VBK	1	344		126	-	99.76		8.31	

Student club	Rounds	Seconds/Round	Dist (km)	Speed km/h
HILOK	1030	42	298,70	24,89
VTK	1028	42	298.12	24.84
VLK	841	51	243.89	20.32
Wetenschappen and VLAK	819	53	237,51	19.79
VGK	810	53	234.90	19.58
Hermes and LILA	793	54	229.97	19.16
НК	771	56	223.59	18.63
VRG	764	57	221.56	18.46
VEK	757	57	219.53	18.29
VPPK	689	63	199.81	16.65
SK	647	67	187.63	15.64
Zeus WPI	567	76	164.43	13.70
VBK	344	126	99.76	8.31

Note that the table above is created using the syntax in Listing 3.10 and is not an image that was included in the document. In fact, all of the examples shown from here on out, until the end of this chapter, are created using the DSLs demonstrated in said examples.

First of all, we note that it does not matter how many spaces are placed inside the elements of the table. The DSL for this table uses of the symbol method |, a nested method - and a new NLHandler.

The first | method is defined in TextContainer, so that it can be used in Chapters as well as Documents (they are both subclasses of TextContainer). It is shown below

Listing 3.11: TextContainer.no

```
/**
2  * Creates a new Table and adds a new TableRow to it.
3  * The Table is added to this.
4  *
5  * @param text The text to use in the first element of the TableRow
6  * @return The created TableRow
7  */
8 TableRow | (Text text) {
9    new TableRow(text) row;
10    new Table(row) table;
11    addContent(table);
```

```
12
13 return row;
14 }
```

The TableRow that is returned form the the | method is a row that is still under construction. By continuously calling | we are able to construct the entire row. This is the reason why we mentioned that symbol methods could be called after another symbol method in Subsection 2.4.1.

The | method that builds the TableRow is defined in TableRow and shown below. We also show the | method that has no arguments, it defines the end of a row.

Listing 3.12: TableRow.no

```
TableRow | (Text text) {
    elements.add(new TableElement(text));
    return this;
}

TableRow | () {
    fte = new FinalTableElement();
    return this;
}
```

The newline method defined in TableRow returns a TableNLHandler.

This handler defines the – nested method and the | symbol method. The former creates a horizontal line in the document and disregards the nesting level. This level could however be used to determine the width of the table for example. The latter is used to start the construction of a new TableRow.

The code for the - and | methods defined in TableNLHandler is shown below.

Listing 3.13: TableNLHandler.no

```
/**
2  * Creates a horizontal line and appends it to our parent
3  *
4  * @param level The level of this nested operator, it is ignored
5  * @return The appended Hline
6  */
7  nested Hline -(Integer level) {
8    return parent().appendHline();
9 }
```

Finally we can write an empty newline to return control of newlines to the enclosing Content

The libraries that are discussed in the next sections are domain specific and show how versatile Neio can be.

#### 3.2.5 Red black trees

Red black trees are trees that are constantly update according to a defined algorithm. When we want to show a red black tree in LATEX, we can typeset an instance of such a tree using for example TikZ. We could also create it in a third party program and include an image of the instance in our document, this solution also works for Markdown.

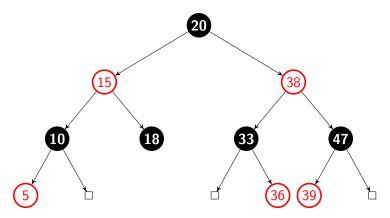
However, since an algorithm is used, and we have a programming model, could not we just build the tree and then display it? Yes we can! All we have to do is take some code that can generate red black trees and add a display method to it. We show an example of the red black trees in use below.

Listing 3.14: rbtDocument.no

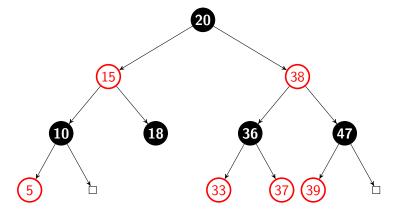
```
1
  [Document]
2
3
       List<Integer> tree = new ArrayList<Integer>();
       tree.add(33);
4
       tree.add(15);
5
       tree.add(10);
6
       tree.add(5);
       tree.add(20);
8
9
       tree.add(18);
       tree.add(47);
10
       tree.add(38);
11
12
       tree.add(36);
       tree.add(39);
13
14
       String numbers = ',',',' + tree.get(0);
15
       for (int i = 1; i < tree.size(); i = i + 1) {
```

```
17
            numbers = numbers + ''', ''', ''' + tree.get(i);
18
  }
19
20
  Given the red black tree of {numbers}
21
^{22}
       RedBlackTree rbt = new RedBlackTree().insert(tree);
23
       return rbt;
24
  }
25
26
  Add 37
27
28
       RedBlackTree rbt2 = rbt.insert(37);
^{29}
30
       return rbt2;
  }
31
32
  Add 51
33
34
  {
       RedBlackTree rbt3 = rbt2.insert(51);
35
       return rbt3;
36
37
  }
```

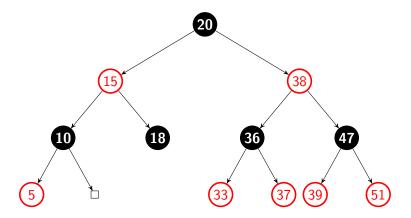
Given the red black tree of 33, 15, 10, 5, 20, 18, 47, 38, 36, 39



**Add** 37



Add 51



This code allows us to very easily create and show red black trees. Of course, we are not limited to red black trees. This can be done for any tree or graph as long as we have code for it and write a method that can display it.

To compare our method to typesetting every instance of the tree in LATEX, we show the LATEX code for the first instance of the tree. This code is not that easy to read, hard to write and prune to syntax errors. For example, this code is newline sensitive, placing empty newlines between the entries throws a syntax error at compilation. The code for the next instance of the tree is very similar, but it is hard to reuse this code.

Listing 3.15: rbtLatex.tex

```
\begin{tikzpicture}[->,>=stealth',level/.style={sibling distance = 5cm/#1,
      level distance = 1.5cm}]
   \node [arn_n] {20}
   child { node [arn_r]
                          {15}
3
   child { node
                 [arn_n]
                          {10}
   child { node [arn_r] {5}
   child { node [arn_x] {}}
7
   child { node [arn_n] {18}
9
10
11
   child { node
12
                 [arn_r]
                          {38}
   child { node
                          {33}
13
                 [arn_n]
   child { node [arn_x]
                          {}}
14
15
   child { node [arn_r] {36}
16
17
   child { node [arn_n] {47}
18
19
   child { node [arn_r] {39}
20
   child { node [arn_x] {}}
^{21}
22
  }
23
^{24}
```

```
25 \ \end{tikzpicture}
```

Note that this tree is not implemented optimally, when we append a new value to the tree in our implementation, we completely clone the tree and add a new node to this new tree. The red black tree in this example also only works for integers.

The reason we clone the tree every time we insert a new value is because otherwise the three trees would be represented by the same object. They would thus all show the latest version of the tree. This problem is discussed in Chapter 5.

### 3.2.6 MetaUML

The next example has already been used in Subsection 3.2.3. Every UML diagram in this document is created using the UML library discussed in this section.

It is also a little different from the other libraries we have seen thus far as it does not directly translate into LATEX. Instead we are making use of MetaUML [19]. How this works exactly is shown in Chapter 4. For now it is enough to know that the object model build by this library is output as a string that is understood by MetaUML. MetaUML then creates a MetaPost file [37] which is included in the document.

In this section a high-level overview of how we create UML diagrams is given. We start out with an example.

```
RedBlackTree

RedBlackTree(): RedBlackTree
RedBlackTree(root: RedBlackNode): RedBlackTree
root(): RedBlackNode
insert(items: List<Integer>): RedBlackTree
insert(value: Integer): RedBlackTree
insert(value: Integer): RedBlackTree
insert(value: Integer, current: RedBlackNode): void
fixTree(node: RedBlackNode): void
fixRedParentUncle(node: RedBlackNode): void
fixRPBUR(node: RedBlackNode): void
fixRPBUR(node: RedBlackNode): void
rotateLeft(node: RedBlackNode): void
rotateRight(node: RedBlackNode): void
preTex(): void
toTex(): String
```

```
RedBlackNode
  RedBlackNode(color: Integer, value: Integer): RedBlackNode
  setColor(color: Integer): void
setParent(parent: RedBlackNode): void
  setRight(right: RedBlackNode): void
setLeft(left: RedBlackNode): void
cloneSelf(): RedBlackNode
grandparent(): RedBlackNode
  uncle(): RedBlackNode
color(): Integer
  value(): Integer
isLeaf(): Boolean
  isRed(): Boolean
  isBlack(): Boolean
  left(): RedBlackNode
right(): RedBlackNode
  parent(): RedBlackNode
  hasLeft(): Boolean
  hasRight(): Boolean
■ isRoot(): Boolean
preTex(): void
  toTex(): String
```

This code reads the project.xml file, that tells it what project we want to create an UML diagram for. The second parameter is the namespace the UML diagram should be created for. This means that the UML diagram in this example is created from the actual classes that make up this document.

When the document is compiled, the neio code files are read and

It is also possible to only show a few classes, or to also show private and protected members.

This is shown in the example beneath.

Uml

Uml(projectXml: String, ns: String): Uml
gatherClasses(): void
show(show: List<String>): Uml
showPrivate(): Uml
showProtected(): Uml
showAll(): Uml
scale(scale: Integer): Uml
preTex(): void
toTex(): String
toMetaUML(): String
id(o: Object): String

UmlClass

UmlClass(type: Type, showPrivate: Boolean, showProtected: Boolean): UmlClass

gatherVars(): void

gatherMethods(): void

type(): Type

name(): String

toMetaUML(): String

In this example the UML diagram shows two of the **actual** classes that were used to create this UML diagram. The code to create these UML diagrams thus reads and displays itself. How this works exactly is explained in Chapter 4.

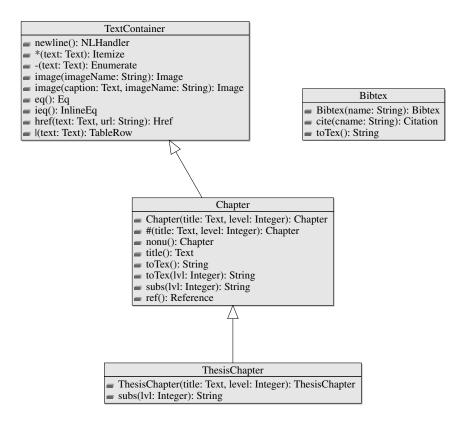
As the code for this library is over 600 lines long, we do not show all of the code.

The problem with the MetaUML tool is that positioning is not done automatically. The way we implemented positioning is like a tree.

We build a tree of levels, the first level contains all the classes that have no super class to be shown as well as the highest level classes. Subclasses are shown on the level beneath their super class.

For example, ThesisChapter is a subclass of Chapter, which is a subclass of TextContainer. If we now show these three classes, they are displayed one under the other because we built a tree with three levels.

To visualize how the levels are being used, we show this example below and also add the Bibtex class to the UML diagram, to show what happens to classes that have no connection to the other classes.



Note that our implementation of UML diagrams only shows what it is asked to show (hence there are no uses arrows) and only draws inheritance arrows for what is shown. These arrows are only shown for direct super classes. This is done to avoid clutter.

### 3.2.7 Chemfig

In this example we approach a different domain, the domain of biology and chemistry. For this domain, we created a library and DSL to create structural formulas. An example of the structural formula for glucose is given below.

```
1  [Document]
2  {
3     new Structure()
4  }
5     $ C * 6 _ OH ^ - C _ ^ OH - O - C _ ^ CH2OH - C _ HO ^ - C _ ^ OH -
7  | Glucose
```

This example produces the following output

The first thing we see, is that we create a new Structure. This provides a newline handler that allows us to instantiate structure formulas using the \$ method and captions for them using the | method. The \$ method shown below, creates an atom and adds it to the structure, the rest of the structural formula is then build by chaining atoms together.

Listing 3.16: StructureNLHandler.no

```
Atom $(Text text) {
    Atom a = new Atom(text.realText());
    a.setParent(parent());
    parent().addAtom(a);
    return a;
}
```

The \* method, followed by a number, defines that the formula is a ring. The number defines the number of edges of this ring. The – method creates a single link from the previous atom to the next, = creates a double link. Finally, the ^ method defines that the next atom is above the previous one, while \_ puts the next atom below the previous one. The code for one of these is found below.

Listing 3.17: Atom.no

```
Atom =() {
1
      return createNext("", "=");
2
  }
3
4
5
  private Atom createNext(String atomText, String link) {
      Atom a = new Atom(atomText);
6
      a.setParent(structure);
      next = a.setLink(link);
8
      return a;
  }
10
```

The toTex method returns a LATEX representation of the structure that is understood by the LATEX package chemfig [5].

### 3.2.8 Lilypond

For the last example we create sheet music. An example musical score is shown below.

Listing 3.18: An example musical score.



In this DSL, a class has been made to represent a note and a method is available to create every note. For example, to create the note C, you call the method c(). The integer that can be passed to a note, lower or heightens the octave of the note.

Every operation that would modify the current Score, first copies the Score to make sure that we do not edit any of the already placed scores.

The code to add the C note to a Score is given below.

Listing 3.19: Score.no

```
public static final Integer C = 0;
/**

  * Copies this Score and adds a new note to the copy

  *
  * @param note   The note to add
  * @param octave   The octave of the note
  * @return A copy of this Score with a new note added to it
  */
Score add(Integer note, Integer octave) {
```

```
10
       return new Score(this, new Note(note, octave, this));
  }
11
12
      Adds the c node to a new Score
13
14
      @param octave The octave of this c
15
      Oreturn A new Score containing a new c
16
17
  Score c(Integer octave) {
18
19
       return add(C, octave);
20
```

We can now query the notes of this score, for example {s.get(8)} gives us the eighth note. Doing this for our example tells us that the eighth node is e.

It is also possible to reverse a score or shift it up or down using the reverse and shift methods respectively. The sequence of notes can also be printed using the print method. These methods are demonstrated in the example below.



This score is read as c, d, e, f, g, a, b.



We can shift every note by one, doing so yields the following notes: d, e, f, g, a, b, c.



Reversing the score results in the following sequence of nodes: b, a, g, f, e, d, c

The code for this example as well as the implementation of the reverse and shift methods is given below.

Listing 3.20: Score.no

```
* Copies this Score and shifts every
2
       node up or down by {@code shift}
3
      Oparam shift The amount to shift the
4
        nodes by
      Oreturn A new Score where every node
5
         is shifted by {@code shift}
6
   Score shift(Integer shift) {
       Score s = new Score(this);
8
9
       List < Note > n = s.notes();
       for (int i = 0; i < n.size(); i = i
10
            + 1) {
11
           Note note = n.get(i);
           note.shift(shift);
12
13
       return s;
14
  }
15
16
17
   * Creates a copy of this Score with
18
       the notes in reverse order
19
     Oreturn A new Scope with the notes
20
       of this Score in reverse order
21
22
  Score reverse() {
       Score s = new Score(this);
23
^{24}
       Collections.reverse(s.notes());
       return s:
25
  }
26
```

Listing 3.21: The code of the example.

```
{
       Score ss = new Score().c().d().e().
2
          f().g().a().b();
       return ss;
  }
4
  This score is read as {ss.print()}.
6
7
       Score sss = ss.shift(1);
8
       return sss;
  }
10
11
  We can shift every note by one, doing
12
      so yields the following notes: {sss
      .print()}.
  {
13
14
       Score s4 = ss.reverse();
       return s4;
15
  }
16
17
18
  Reversing the score results in the
      following sequence of nodes: {s4.
      print()}
```

To typeset these music sheets, the object model is written out in a format that is understood by LilyPond [29]. This is written to a file and the file is passed to the lilypond command line tool. This tool generates a PDF that is then inserted as an image in the document.

IMPLEMENTATION 77

# Chapter 4

# Implementation

to use Neio, we also had to build a compiler. How we implemented it, as well as some lower level concepts used in Neio code files, is explained in this chapter.

## 4.1 Compile flow

We start out by showing a diagram that depicts the flow a source file from Neio source up to to the final LATEX file.

4.1 Compile flow 78

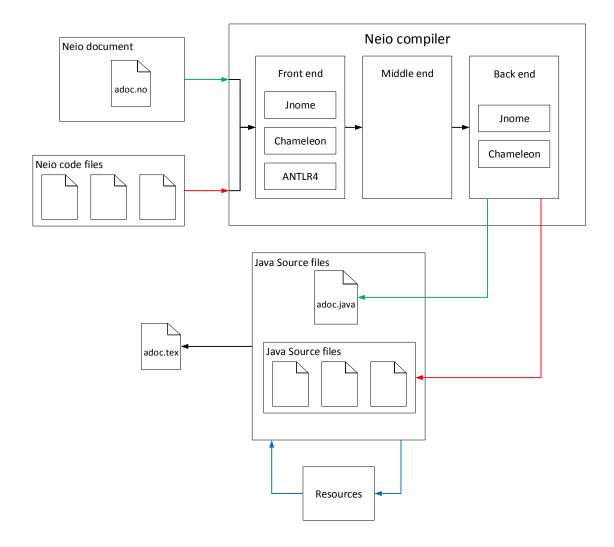


Figure 4.1: Illustration of how a Neio document is compiled. Neio code files translate to Java classes, Neio script files to a Java file with a main method. This then gets output to TeX using the Neio standard library.

First of the Java and Neio library (the Neio code files) are read by the front end of the Neio compiler and translated into an AST. Then Neio documents are read by the front end and it is also transformed in an AST. The latter AST is past to the middle end of the compiler and is transformed to an object model that can be output to valid Java. Then the transformed AST and the ASTs for the Neio library are passed to the compiler back end which generates the final Java code. This final Java code is then be compiled and executed again. During the execution, resources might be created, if so, they can be used in the next step. This next

4.1 Compile flow 79

step is the building of the document to a final representation. The way it always happens at is the time is as follows. The TexFileWriter is invoked with the root of the AST of the Neio document, this then calls preTex and toTex on this root. preTex allows a document to prepare itself for being printed to LATEX. A common preparation is to add a package that is uses to the root document. toTex does what is says and just creates a LATEX representation of the object. Once that is completed, it writes out this TeX string to a file named the same as the Neio document but with the TeX extension. Finally the path to this new TeX file is passed to the TexToPDFBuilder which compiles the LaTeX code. The code for these two files is shown below.

Listing 4.1: TexFileWriter.no

```
namespace neio.io;
1
2
3
  import java.nio.file.*;
  import java.util.Arrays;
4
  import neio.lang.Content;
   * Recursively converts Content into a TeX string and prints it to a file
8
9
  class TexFileWriter implements Writer;
10
11
  Content content;
12
13
14
   * Initializes the writer with a root content
15
16
     Oparam content The root Content
17
18
  TexFileWriter(Content content) {
19
       this.content = content;
20
  }
21
22
23
   * Creates the TeX string
24
25
   * @param name The name of the file to print to (without extension)
26
27
   * Oreturn The path to the created file
28
  String write(String name) {
29
       content.preTex();
30
       String toWrite = "% Name: " + name + "\n" + content.toTex();
31
       Path path = Paths.get(name + ".tex");
32
       Files.write(path, Arrays.asList(toWrite.split("\n")));
33
       System.out.println("Wrote " + path.toAbsolutePath());
34
35
       return path.toAbsolutePath().toString();
36
37
```

Listing 4.2: TexToPDFBuilder.no

```
namespace neio.io;
1
2
3
  import java.io.File;
   import java.util.List;
  import java.util.ArrayList;
7
8
   * @author Titouan Vervack
9
10
  class TexToPDFBuilder implements Builder;
11
12
   * Compiles a TeX file
13
14
   * @param name The name of the file to compile (without extension)
15
16
17
   void build(String name) {
       if (name != null) {
18
           List<String> command = new ArrayList<String>();
19
           command.add("latexmk");
20
           command.add("-pdf");
command.add("-dvi-")
21
22
           command.add("-bibtex");
23
           command.add("-lualatex");
^{24}
           // Make sure the delimiters are correct
25
           command.add(new File(name).getAbsolutePath());
26
27
           ProcessBuilder pb = new ProcessBuilder(command);
28
           pb.redirectOutput(ProcessBuilder.Redirect.INHERIT);
29
           pb.redirectError(ProcessBuilder.Redirect.INHERIT);
30
           pb.start().waitFor();
31
32
           String texlessName = name.substring(0, name.length() - 4);
33
           name = texlessName + ".pdf";
34
           System.out.println("Wrote " + name);
35
36
37
           command.clear();
           command.add("latexmk");
38
           command.add("-c");
39
           command.add(name);
40
41
           ProcessBuilder pb2 = new ProcessBuilder(command);
42
43
           pb2.redirectOutput(ProcessBuilder.Redirect.INHERIT);
           pb2.redirectError(ProcessBuilder.Redirect.INHERIT);
44
45
           pb2.start();
           System.out.println("Cleaned files for " + texlessName);
46
       }
47
  }
48
```

### 4.2 Used libraries and frameworks

to build the compiler, we made use of three open-source libraries and frameworks. They are explained in the subsections below.

### 4.2.1 ANTLR4

The ANTLR4 [1] library is an open-source parser generator released under the BSD license. It is used to parse all of the Neio files, the Neio documents and the Neio code files.

We used ANTLR4 in the following way. First we created two grammars, consisting of a lexer and a parser, for the Neio documents and classes, thus four files in total. These are written using the ANTLR4 DSL. These grammars are then fed to the antlr4 command line tool.

This tool generates Java classes using the grammars you defined. These classes allow us to visit the Concrete Syntax Tree of a file using the visitor pattern.

Whilst visiting every rule and terminal, we build an Abstract Syntax Tree (AST) for the given file. This can then be manipulate in later stages.

The objects that are used in the AST are objects defined in Jnome and Chameleon. These object, in combination with the ANTLR4 grammars and visitor classes, thus form the front end of our compiler.

#### 4.2.2 Chameleon and Jnome

Chameleon and Jnome are both projects that were developed by professor van Dooren and released under the MIT license.

Chameleon [2] is a framework that allows you to model a programming language, or as we have shown in this thesis, a markup language. It does so by providing objects that represent concepts commonly used concepts of programming languages. A few examples of such objects are the following: Expression, VariableDeclaration, Type, MethodInvocation, InfixOperatorInvocation,....

Jnome [10] uses Chameleon and implements the front end and back end of a compiler specifically for the Java language. It is a front end for the Java language because it can read Java and build an AST from it (using ANTLR3). To do this it extends the object model available in Chameleon with Java specific objects.

It is also able to output Java code, given an AST built out of Jnome and Chameleon objects. This part is the back end of the Neio compiler.

Java library. This is needed because our objects use Java objects, such as Strings. It reads all of the source files and libraries that are specified in a file called project.xml. This file was mentioned earlier in Subsection 3.2.6. This same file is used to read Neio projects in our compiler. The project.xml for the Neio compiler is shown below.

Listing 4.3: project.xml

```
project name="Thesis">
       <language name="neio" />
2
3
       <baselibraries>
           library language="Neio" load="true" />
4
       </baselibraries>
5
6
       <sourcepath>
           <source root="../../examples/0.8/lib" />
7
       </sourcepath>
       <br/>
<br/>
dinarypath>
9
            <jar file="../build/libs/chameleon.jar" />
           <jar file="../build/libs/jnome.jar"</pre>
11
            <jar file="../build/libs/neio.jar" />
12
       </binarypath>
13
  </project>
14
```

Having obtained an AST in Subsection 4.2.1, we now manipulate it in the middle end of our compiler. The transformed AST is then output to Java using the Jnome back end.

### 4.3 Translation

### 4.3.1 Reasons for choosing Java

The reason we chose for Java was multi-fold. First and foremost, we already had a front- and back end available for Java thus this saved us a lot of time. A second reason is because Java is platform independent which does not restrict us to a single operating system. Java is also very well known in the current world of computer science, which is why we chose to base our programming model on it. Since the Neio semantics now resemble the Java semantics so well, the reasons for going with Java are enforced.

#### 4.3.2 Fluent interface

We tried to make use of fluent interfaces as much as possible, in the design of the language, by using method chains. But also in the libraries we implemented in Section 3.2. We chose to this as it imposes less boilerplate code (no constant variable declarations for example) and because it is clearer to read what we are doing. Instead of passing a ton of arguments to one method, we split it up in multiple methods and build up one object.

### 4.3.3 Outputting

Neio makes use of Context Types, which are not something that exists in other languages as of yet. This means we had to create a custom translation for it. The way we choose to translate it, is by breaking up the method chains created in the Text mode of a Neio document at every method call. A variable is then assigned to the output of this method call.

To know what to call our method on, ContextTypes were implemented as an extension of RegularJavaTypes in Jnome. A RegularJavaType is what we know as a Java class. This ContextType adds a virtual inheritance relation between itself and the actual type of the method, and one between itself and the ContextType of this method. The latter ContextType represents the Context at that point as we saw in Section 2.5.

When a lookup reaches a ContextType and asks us what our inherited members are (the members you normally receive from super classes), we return the members of all our inheritance relations. We only have two inheritance relations, the actual type and the ContextType. If what we are looking for was somewhere we could see, we return it and we are able to make use of it. Lookups are provided by Chameleon and explaining how exactly they work is out of the scope of this Thesis.

We also hold a Stack of all the variables that have already been declared. When a lookup for a method, described above, succeeds, we check what type is returned by this method and look for the first variable in our stack of that type, popping everything on the way there. We then know what to call our method on.

Lastly, we also have to translate this. The translation is analogue to the previous step, where we searched what to call the method on. To have a starting point, we just replace this with the last defined variable and let the ContextTypes do their job. A translation for the very first example we looked at in this document is shown below.

Listing 4.4: testInput.no

```
new Document()
  [Document]
                                                     .newline()
                                              2
 # Chapter 1
2
                                              3
                                                     .newline()
 ## Chapter 2
                                                     .#("Chapter 1")
                                              4
4
 # Chapter 2
                                                     .newline()
                                                     .text("This is some text in the
  This document contains {nearestAncestor
                                                        first Paragraph");
      (Document.class).directDescendants(
      Chapter.class).size()} top-level
      chapters.
```

Listing 4.5: testInput.java

```
package testInput;
  import neio.fib.*;
3
  import neio.io.*;
  import neio.lang.*;
  import neio.stdlib.*;
  import neio.thesis.*;
  import neio.stdlib.chem.*;
  import neio.stdlib.graph.*;
10
  import neio.stdlib.math.*;
11
  import neio.stdlib.music.*;
  import neio.stdlib.uml.*;
  import java.util.*;
14
15
  public class testInput {
       public static void main(String[] args) {
16
           Document input = new Document();
^{17}
           createDocument(input);
18
           finishDocument(input);
19
       }
20
21
       public static void createDocument(Document input) {
22
           NLHandler $var0 = input.newline();
23
           Chapter $var1 = input.hash(new Text("Chapter 1 "));
24
           NLHandler $var2 = $var1.newline();
25
           Chapter $var3 = $var1.hash(new Text("Chapter 2"), 2);
26
           NLHandler $var4 = $var3.newline();
27
           Chapter $var5 = input.hash(new Text("Chapter 2"));
28
           NLHandler $var6 = $var5.newline();
29
           NLHandler $var7 = $var6.newline();
           Paragraph $var8 = $var7.text(new Text("This document contains ").text(
31
               new Text(("") + ($var5.nearestAncestor(Document.class).
               directDescendants(Chapter.class).size()))).text(new Text(" top-
               level chapters.")));
       }
32
33
```

```
public static void finishDocument(Document input) {
    java.lang.String $var0 = new TexFileWriter(input).write("testInput");
    new TexToPDFBuilder().build($var0);
}
```

Note that the middle end has created three methods. The first one is the main method, it allows the document to be executed as a whole. It initialises the document class and then calls createDocument which actually creates an object model that represents the document. Finally finishDocument compiles the document to TeX and create a PDF of of it.

The reason we split up this process is so that we can call the creation of the document separately. This allows us to include a document into another document by just calling the createDocument method with an appropriate argument.

The middle end also imported every namespace in the Neio library as well as the java.util namespace as those might be needed. The name of the Java class that is generated is the same as the name of the Neio document.

It is important to note that we only split up the methods chains created in Text mode. The code written in code mode is kept as is, only occurrences of this are replaced.

### 4.3.4 Reflection

In Subsection 2.6.1 we touched on addClassMapping and said that the class mapping was comparable to a factory. This is true and the need for it appeared when we wanted to create drop in replacements for objects that are created in Neio Text mode. For example, if we want to use a special kind of Chapter in the Document document class, we would have to create a new document class that overrides the # method, we also have to override the nested # method in Chapter because otherwise only the first Chapter would be an instance of the special Chapter.

This is a problem that is normally solved by using factories. When you know your object can create instances of other objects, in our example a Document creates Chapters, that might be changed with more specialized forms of that object, you create a factory. This factory

however would have to be generic enough to be used almost anywhere. Even if we are able to do this, the writer of the specialized object would still have to create a new factory. Factories usually are also very similar to each other, yet the code can not be reused.

For this reason we chose to swap out new calls with calls that use reflection to initialize the object. This is a process that is executed in the middle end of the compiler. Only the new calls of objects extending Content have been replaced as those should be the only thing a user wants to replace. Every new call gets replaced by the getInstance call in the Content class. The code needed to understand this process, as well as an example call is given below.

Listing 4.6: Content.no

```
// Determines as what class new Content instances will be instantiated
2
  private final Map < Class, Class > _classMapping;
      return descendants;
3
  }
4
5
6
   * Returns the Class as which a Class should be instantiated
7
   * Oparam klass The key of which we want to get a mapping
9
   * Oreturn The mapping found for key {Ocode klass}
10
11
  final Class classMapping(Class klass) {
12
      return _classMapping.get(klass);
13
14
15
16
   * Overrides default behaviour and instantiates new instances of {@code
17
       oldClass } as instances of {@code newClass}
18
   * @param oldClass The old class object
19
   * @param newClass The new class object
20
   * /
21
  public final <T> void addClassMapping(Class<T> oldClass, Class<? extends T>
22
      newClass) {
       _classMapping.put(oldClass, newClass);
23
  }
^{24}
25
26
   * Used instead of a new call. This allows to create more specific objects
27
       than originally
   * specified. e.g. this allows to create SpecialChapter's instead of Chapter's
28
29
   * It will create an instance of {@code klass} in case the class mapping has
30
       not been overridden.
   st If the mapping was overridden, an instance of the overriding class will be
31
       created.
32
   * Oparam klass
                        The class to instantiate
33
   * Cparam paramTypes The types of the parameters
                        The parameters to instantiate {@code klass}
   * @param params
35
   * Oreturn The instantiated object
37
```

```
38 | public final <T> T getInstance(Class<T> klass, Class[] paramTypes, Object[]
      params) {
       // Have I been overridden?
39
       Class result = classMapping(klass);
40
       Content current = parent();
41
       // Has any of the parents been overridden?
42
       while ((current != null) && (result == null)) {
43
           result = current.classMapping(klass);
44
           current = current.parent();
45
       }
46
47
       // No parents have been overridden
48
       if (result == null) {
49
           return instantiate(klass, paramTypes, params);
50
51
       // A parent was overridden, use that type
52
53
       else {
           return (T) instantiate(result, paramTypes, params);
54
55
  }
56
57
58
   * Does the actual instantiation of an object using Reflection
59
60
    * Oparam klass
                         The class to instantiate
61
    * @param paramTypes The types of the parameters
62
63
    * @param params
                         The parameters to instantiate {@code klass}
    * @return The instantiated object
64
65
```

```
Chapter chapter = getInstance(Chapter.class, new Class[]{Text.class, Integer.
      class}, new Object[]{title, 1});
```

If you already understood the code from the comments, this paragraph can be skipped. Every instance of content has a mapping from Class to Class, by default this mapping is empty. When we want to use a specialized form of an object we call addClassMapping, as we did in Subsection 2.6.1. Take the following example.

We added a mapping from Enumerate.class to FibEnumerate.class in the root Document. When we now call getInstance (Enumerate.class, ...), the root of the call is Chapter, not Document. But fear not, class mappings are looked up recursively. In the getInstance method we first check if we have a mapping for the class we are trying to instantiate. If we do the value of the mapping is used, if not we recursively check the mappings of our parents. If no mapping has been found after this, we just instantiate the class that has been given as

argument to getInstance.

### 4.3.5 Escaping

A Neio document is parsed and converted a number of times. First, it is parsed by the Neio compiler, then it is transformed into Java code that is then converted into IATEX code (in most cases). This means that we have to handle escaping of characters in three different languages. Escaping in Neio is easy, just add a \ before the character you want to escape and it is escaped. Of course \[ntr]\] hold on to their special meaning. However \[bfu]\] have not been included in Neio as they are not yet needed. In the future it might prove useful to include these special characters to.

Moving on, we can not just translate these characters straight to Java as we have no special meaning for \[bfu]. For this reason the middle end transforms every backslash into two backslashes, and transforms a double backslash (used to denote an escaped single backslash in Neio) to four backslashes in Java as this is how you represent an escaped backslash in a Java String. We then do some further translation in the Text class to assure that we are able to create valid latex. LATEX only has ten special characters thus replacing is quite easy. The code to do the replacing is shown below.

Listing 4.7: Text.no

```
1
2
    * Creates the TeX representation of this Text
3
    * Printing out {@code realText} or the TeX representation of {@code text}
    * Replaces special characters to create valid TeX
5
6
    * Oreturn The TeX representation of this Text
7
    */
   String thisToTex() {
9
        String me = "";
10
        if (text == null) {
11
             // Create valid TeX text
12
             me = realText.replaceAll(BS + BS, ESC + "textbackslash{}"); me = me.replaceAll(BS + "([^#$%&^{\}" + BS + "])", "$1");
13
14
             me = me.replaceAll(ESC, BS);
15
             me = me.replaceAll(BS + "~", BS + "textasciitilde{}");
me = me.replaceAll(BS + "\\^", BS + "textasciicircum{}");
16
17
        } else {
18
             me = text.toTex();
19
        }
20
```

return me;

### 4.4 Resource usage and creation

### 4.5 Limitations

Whilst development of the compiler a few limitations were discovered, they are discussed in this section.

### 4.5.1 Windows

In Section 4.4 we saw that command line commands are sometimes executed. Executing commands using the Java Process class, however came with some not immediately visible issues. When developing on Linux, executing shell commands from Java worked perfectly fine and gave no problems whatsoever. However, on Windows some of the processes would hang, some would always hang, some would hang only once in a while. The reason for this is that Windows offers only a limited buffer size for the input and output streams of a process, the program deadlocks if the input is not written or read. After closer inspection of the Javadoc [8] this became apparent but since the API for the Process class is so simple, it was a surprise nevertheless. The fact that we have to worry about different platforms on a cross-platform language as Java is also something that is not too common.

### 4.5.2 Back end

About every back end has some limitations on how we can translate our Neio document. This is because Neio can use symbols as methods. In Java, methods can only contain letters, numbers, \$ or \_. This means we have to translate the symbol method to valid Java identifiers. The way we do this is very straightforward. We use the following mapping and just replace any occurrence of a key in a method or method invocation with the value.

4.5 Limitations 90

```
"#" -> "hash"
2    "*" -> "star"
3    "=" -> "equalSign"
4    "^" -> "caret"
5    "-" -> "dash"
6    "_" -> "underscore"
7    "(" -> "backquote"
8    "$" -> "dollar"
9    "|" -> "pipe"
```

With other back ends we would do the same, some languages allow you to overwrite operators but doing this might lead to unexpected side effects, thus just translating it this way seemed like the safest way to go. FUTURE WORK 91

## Chapter 5

## Future work

As mentioned before in Section 2.9 there are still a few things that could be done in the future. There are however a lot more things that could be done and we discuss them in the rest of this chapter.

### 5.1 Automatisation

### Automatic StringBuilder

Since we are constantly translating to TeX and other formats, we use StringBuilder very often. An example is almost every toTex method, we show the one defined in Content as an example.

Listing 5.1: Content.no

5.1 Automatisation 92

```
StringBuilder tex = new StringBuilder();
for (int i = 0; i < content.size(); i = i + 1) {
    tex.append("\n").append(content.get(i).toTex());
}</pre>
```

In fact we use it so many that in the future it might be better to just replace Strings by StringBuilders at all times. This would make it easier for people to write packages. The performance gain, or loss of this would have to be investigated though.

### Automatic return of the object itself

Another pattern that is often encountered is that we often return the same object even when we have not changed it. This is to make use of fluent interfaces, as said before. An example from the InlineEq class is shown below.

Listing 5.2: InlineEq.no

```
* Creates a square root and adds it to this equation, returning this to allow
        further chaining
3
4
     Oparam root The base of the square root
   * Oparam arg The value to take the square root of
5
   * Oreturn This equation
7
  InlineEq sqrt(Content root, Content arg) {
8
      content().add(new Sqrt(root, arg));
9
      return this;
10
11
```

We actually already have a solution for this, constructors. Constructors do not tell us in their signature what they are going to return, but we do know what they are going to return. No explicit return is needed either. Implementing this for other methods would allow us to lower the amount of boilerplate code that has to be written.

### **Automatic Text conversion**

The final automation is needed to counter what we saw with the math library. There we had to open an inline code block just to open Text mode because that creates a Text. A

5.2 Moving content 93

way to prevent this would be to automatically convert Content, or maybe even Object in a generic way. In Java this is done through the toString method while in python the repr and str methods are used for this. However if we were to implement it as naively as this, we would lose any kind of static typing when a Text is expected as an argument. This does not seem like a good thing, especially when you consider that not every object might have a real textual representation. This means that automatically transforming an object to text might in a lot of cases not even create a sensible result.

A better way might be to define an interface, Representable, that has a repr method. Any class that implements this class, and the native types such as String and Integer, could than be transformed into a Text using this repr method.

## 5.2 Moving content

When we assign content to a variable, and apply small changes to it later on, it is possible that the content moves or that both the original and the new content are changed. This is because the language does not enforce every placed content to be static and in turn this is the responsibility of the library developer and the user of those libraries. We showed that we had to make sure that we cloned our objects when making changes to them before in Subsection 3.2.5 and Subsection 3.2.8. It would be a very nice feature if Neio could enforce this behaviour by itself and take care of the implementation of clone by itself. A way that this might be implemented is by copying objects that are going to be changed through reflection.

5.3 Use 94

### 5.3 Use

### 5.4 Packages

### Package support

No package system was provided in this thesis, but in the future however it would be beneficial to do so. It would allow users without a lot of technical expertise to import new functionality easily. The way packaging is handled right now is straight through IATEX packages and Java packages. We allow you to add IATEX packages through the addPackage method in Document and we allow you to import Java classes through the regular import namespace statements. To actually include a Neio library in the document, right now the source code for it has to be put into a folder, next to the rest of the standard Neio library. In the future we should be able to download (manually or automatically) packaged libraries and have them be stored somewhere else.

### Implementation of packages

As we said before, non of the document classes or libraries that were created in this thesis are complete, and we also did not create that many of them, in comparison to what Java and LATEX offer that is. In the future a lot of work should be put into writing complete libraries and porting libraries from other available languages such as LATEX. This can however be done by the community for a big part, as is what happens for Java and LATEX.

## 5.5 Compiler improvement

### Efficiency

The compiler is fast enough for small documents but for a large document like this thesis, the compile time starts to ramp up. The thesis takes about 30 seconds to a minute to compile,

depending on the compute power of your computer. This is on top of the latexmk that has to run afterwards and that also takes quite a while. However no attention was paid to the efficiency of the compiler in this thesis, as thus it is likely that there are a lot of optimizations that can be performed.

#### Other front- and back ends

Another compiler improvement that could be created in the future, is that we could create different front and back ends. We do not have to output to Java, if someone were to write a back end for python for example, we could make use of that back end. The same thing is possible for the front end. The syntax of Neio documents and code files could be changed, as long as the fundamental design decisions of the language are kept the same, such as ContextTypes and nested methods, another syntax could be designed. As long as a front end is developed that can parse a Neio document with a different syntax that can still parse the document into the same AST as is used at this moment, it is possible to change the currently used syntax..

### 5.6 Tool improvement

To effectively be able to use Neio, we need some improvements to the tools available right now. Chameleon provides support for the Eclipse IDE and in the future this could be used to improve our tooling.

### Syntax highlighting

One of the base requirements to work with a language is syntax highlighting. This is one of the features that Chameleon provides through its Eclipse integration. However it was not made for a language such as Neio where we represent a document as a series of method chains. At this moment the Eclipse integration treats the entire method chain as a whole instead of every method call separately. This would cause problems as everything between codeblocks

would be colored the same and we would not be able to have different colouring for special methods, such as the symbol methods.

It should however work out of the box for Code mode and the code files.

### Auto completion

A second feature that Chameleon offers and that is very useful when writing in a certain language, is auto completion. This should again work well in Code mode and code files, however it would be a bit harder in Text mode. This is because it would have to know when to auto complete and when it should just let us write. It might be best to hide the auto complete behind a hotkey while in Text mode. This still allows the user to get a list of methods he could use at a certain time, but would remove the annoyance of having a pop-up for auto completion come up at every word that is written.

### IDE with preview

The last thing that would be a very nice to have for a markup language like this, would be an IDE that previews the document. This is not something that is supported by Chameleon and would thus require quite a bit more work. The compiler in its current state is also not ready for this, as said before it gets quite slow on large documents. As it would have to be constantly running in the background, it would use a lot of energy (which is problematic for laptops) and would be too slow to actually produce results in time.

## Chapter 6

## Conclusion and reflection

### **Summary**

In this thesis we researched the current state-of-the-art concerning document creation (Chapter 1). From this research we concluded that a good combination of user-friendliness and customisability was hard to find. To achieve such a combination we created a new markup language, called Neio, that uses a modern, object-oriented programming model.

We decided on the features (Chapter 2) that Neio needed to achieve its goals and used them to create the examples in Chapter 3. These examples were created to show the simplicity and flexibility of Neio.

to prototype Neio, we also created a compiler and a standard library for it (Chapter 4).

Finally, we discussed the work that can and should still be done on Neio in the future in Chapter 5.

### Reflection

We used the language for four months (the rest of the time was spent developing it), wrote a Thesis in it and wrote a complete compiler for it. After all of this we found that it might have been better to just use almost pure Java syntax for the code mode. We changed a few things, such as adding syntactic sugar for new calls and calling packages, namespaces. The latter was done because in our eyes, namespace is a more general term, that better explains what this keyword does.

But by choosing different keywords, we ended up with name conflicts. At some point in the development we wanted to make use of a class in a package called namespace. Of course this did not work out well as the parser did not know how to parse a package name containing a keyword. By changing the syntax as much as we did, it also became impossible to run Javadoc on our code files without making major changes to them.

Disallowing a substantial part of Java (anonymous classes, shorthand operators such as ++ and +=,...) also hurts the adoption rate of the language as you almost always to invest some time converting a Java file into a Neio file.

We think that it would have been better to have just used Java syntax, with the modifications defined in Section 2.4, for the code mode. In the future we might then also add the proposition made in Subsection 5.1.

In text mode it is annoying that we can not use more than one character as a symbol method. For example, it is not possible to use meaningful symbol such as -> as a method. It would thus have been better if we had implemented symbol methods in a way that allows for more than one character to be used as an identifier.

As shown in Subsection 3.1.4, some work should also be put into making symbol methods more flexible. It should be possible to easily pass parameters of an object through symbol methods.

### Conclusion

In the end, we showed that is possible to create a healthy mix between user-friendliness and a powerful programming model. We showed that the language can be as easy to read, write and learn as Markdown. While it can also be as powerful and diverse as LATEX whilst remaining quite legible.

A lot of work can still be done, and should still be done to allow the widespread use of Neio, as is shown in Chapter 5, but we think that good basis has been provided.

BIBLIOGRAPHY 100

# **Bibliography**

- [1] ANTLR4. http://www.antlr.org/. Accessed: 18-05-2016.
- [2] The Chameleon framework. https://github.com/markovandooren/chameleon. Accessed: 18-05-2016.
- [3] Stephen Checkoway, Hovav Shacham, and Eric Rescorla. Are text-only data formats safe? or, use this LaTeX class file to pwn your computer. In *LEET*, 2010.
- [4] Steve Checkoway, Hovav Shacham, and Eric Rescorla. Don't take LaTeX files from strangers. 2011.
- [5] Chemfig LaTeX package. https://www.ctan.org/pkg/chemfig. Accessed: 26-05-2016.
- [6] Fluent interfaces. http://www.martinfowler.com/bliki/FluentInterface.html. Accessed: 27-05-2016.
- [7] GitHub Flavored Markdown. https://help.github.com/enterprise/11.10.340/user/articles/github-flavored-markdown/. Accessed: 27-05-2016.
- [8] Javadoc for process. https://docs.oracle.com/javase/8/docs/api/java/lang/ Process.html. Accessed: 09-05-2016.
- [9] JMathTeX: TeX in Java. http://jmathtex.sourceforge.net/. Accessed: 08-05-2016.
- [10] The Jnome framework. https://github.com/markovandooren/jnome. Accessed: 18-05-2016.
- [11] Donald E. Knuth and Michael F. Plass. Breaking Paragraphs into Lines. Software -Practice and Experience, 11:1119–1184, 1981.

BIBLIOGRAPHY 101

[12] KOMA-Script guide. http://texdoc.net/texmf-dist/doc/latex/koma-script/scrguien.pdf. Accessed: 18-05-2016.

- [13] LaTeX introduction. https://latex-project.org/intro.html. Accessed: 20-05-2016.
- [14] LyX homepage. https://www.lyx.org/. Accessed: 25-05-2016.
- [15] Markdown homepage. https://daringfireball.net/projects/markdown/. Accessed: 07-05-2016.
- [16] Markdown syntax. https://daringfireball.net/projects/markdown/syntax. Accessed: 08-05-2016.
- [17] MediaWiki LaTeX extension. https://www.mediawiki.org/wiki/Extension:Math. Accessed: 08-05-2016.
- [18] Memoir user guide. http://tug.ctan.org/tex-archive/macros/latex/contrib/memoir/memman.pdf. Accessed: 18-05-2016.
- [19] MetaUML. https://github.com/ogheorghies/MetaUML. Accessed: 18-05-2016.
- [20] LaTeX Minipage. http://www.sascha-frank.com/latex-minipage.html. Accessed: 09-05-2016.
- [21] The neio library. https://github.ugent.be/tivervac/neio/tree/master/examples/0.8/lib/neio. Accessed: 27-05-2016.
- [22] Neio source code of this thesis. https://github.ugent.be/tivervac/neio/blob/master/examples/0.8/input/thesis/thesis/thesis.no. Accessed: 27-05-2016.
- [23] Office object insertion. https://support.office.com/en-us/article/Insert-an-object-in-Word-or-Outlook-8fc1ea53-0e01-4603-a4cf-98c49b6ea3f5.

  Accessed: 27-05-2016.
- [24] Pandoc. http://pandoc.org/. Accessed: 08-05-2016.
- [25] Pandoc filter example. http://pandoc.org/scripting.html#json-filters. Accessed: 25-05-2016.
- [26] Till Tantau. The TikZ and PGF Packages.

BIBLIOGRAPHY 102

- [27] The Colemak keyboard layout. http://colemak.com/. Accessed: 27-05-2016.
- [28] The LaTeX beamer package. https://www.ctan.org/pkg/beamer. Accessed: 27-05-2016.
- [29] The LilyPond homepage. http://lilypond.org/. Accessed: 30-05-2016.
- [30] The MediaWiki homepage. https://www.mediawiki.org/. Accessed: 27-05-2016.
- [31] The neio compiler. https://github.ugent.be/tivervac/neio/tree/master/neio. Accessed: 27-05-2016.
- [32] The PanPipe Pandoc filter. http://chriswarbo.net/git/panpipe/index.html. Accessed: 27-05-2016.
- [33] The StackExchange homepage. https://www.stackexchange.com/. Accessed: 27-05-2016.
- [34] The Wikipedia homepage. https://www.wikipedia.org/. Accessed: 27-05-2016.
- [35] The Wiktionary homepage. https://www.wiktionary.org/. Accessed: 27-05-2016.
- [36] W3C EBNF specification. https://www.w3.org/TR/REC-xml/#sec-notation. Accessed: 27-05-2016.
- [37] What is MetaPost. www.tex.ac.uk/FAQ-MP.html. Accessed: 30-05-2016.