



Rheinische
Friedrich-Wilhelms-
Universität Bonn



Institute for Computer Science
Department VI
Autonomous Intelligent Systems

RHEINISCHE FRIEDRICH-WILHELMS-UNIVERSITÄT BONN

MASTER THESIS

Neural Machine Translation

Author:

Arul

Ehsan

First Examiner:

Prof. Dr. John DOE

Second Examiner:

Prof. Dr. John DOE

Advisor:

John DOE

Submitted: June 4, 2017

Abstract

Describe your approach and results shortly in the abstract. The abstract should really already tell the reader what to expect. Do not try to build suspension, this is not a Hollywood movie, it is a (your!) scientific thesis. The abstract is usually the last thing you write, even if it is the first thing you read here.

Contents

1	Introduction	1
2	Language	3
2.1	I or We?	3
2.2	Quotes and Emphasis	3
2.3	Line Noise and Colloquial Expressions	3
3	Comments on Writing in LaTeX	5
3.1	L ^A T _E X's Wide Margins	5
3.2	Paragraphs are Empty Lines	5
3.3	Typesetting Math as Part of Your Text	6
3.4	Referencing Sections, Figures, etc.	6
3.5	Citing Other Works	7
3.6	Floats	7
3.6.1	Figures	7
3.6.2	Tables	8
3.6.3	Algorithms	8
3.7	Miscellaneous	9
3.7.1	Enumerations	9
3.7.2	Front, Main and Back Matters	9
3.8	Compiling The Document	10
3.8.1	Issues With This Template	10
	Appendix	11

List of Figures

3.1	Nyan cats	7
-----	---------------------	---

List of Algorithms

1	How to write algorithms (Small example from the algorithm2e documentation)	9
---	--	---

1 Introduction

It is probably a good idea to split your thesis up into several files, so you can work on them individually. This is an example of such a file, which ends up in the main file but can be compiled by itself without all the other chapters. This will save you tons of time.

2 Language

2.1 I or We?

Use the “we” form, even when you’re writing alone. Imagine it to be the reader and you, who are going through the text together. It gets you closer to your audience.

2.2 Quotes and Emphasis

Note that double ”quotes” on your keyboard are not typeset correctly. English quotes should look like “this”, while German quotes are typeset like „this“. Note the difference. Some editors insert the correct quotes automatically for you.

Use emphasis sparingly. It clutters the text. In general, use *italics*, which is the one standing out the least. In captions, it is sometimes useful to use **bold** for the signal words left and right, top and bottom: There, you want them to pop out.

2.3 Line Noise and Colloquial Expressions

Do not use fill words which do not carry meaning. Try to be as specific as possible. This does not mean as short as possible, it means that you should have something to say when you write something. If you don’t know what you want to say, think again, *then* write.

Don’t use colloquial expressions. In particular, don’t use any of the short forms *don’t*, *aren’t*, *isn’t*, *you’re*, *we’re*, Do not¹ use ellipsis (...) in the text, as it looks like you’re trailing off.

¹this is much better!

3 Comments on Writing in LaTeX

3.1 L^AT_EX's Wide Margins

Many students complain about large L^AT_EX standard margins. Typographers have a rule, though: You shouldn't have more than about 70 characters per line (when using single spacing). Otherwise your eyes have trouble jumping to the beginning of the next line. This limits your column width severely, unless you use a big font. A thin column width results in large margins. This is (aside from portability) the reason why books are smaller than A4 paper, and why newspapers have multiple thin columns.

If you want to toy with margins nevertheless, consider this: When you open your double-sided printed thesis, the white region in the left/right and center region should have equal widths. It follows that outside margins are larger than the margins inside. Consequentially, on a left page, left margins are larger than right (i.e. center) margins, and on a right page, right margins are larger than the left (i.e. center) ones. The margin on the bottom should be larger than on the top. These rules are automatically used by the KomaScript classes (scrartcl, scrbook, scrprpt). Do not meddle with them.

To meddle with the margins, use the DIV option of the `documentclass` command. Larger values result in smaller margins. If you want to bind your thesis, make sure to include some space for it using BCOR. Ask the print shop how much to use for your specific binding needs. Never, ever, use the `geometry` package.

3.2 Paragraphs are Empty Lines

Paragraphs are separated by empty lines in the L^AT_EX code. Never use the double backslash for this purpose. L^AT_EX inserts page breaks preferably between paragraphs. A double backslash lets you stay in the same paragraph, so L^AT_EX does not know where it would be good to insert page breaks, figures, tables, algorithms, and the like.

Many people find the indentation of the first line in paragraph odd. Look into a favorite book or newspaper of yours and you'll see that this way of introducing new

paragraphs is used everywhere. Note that it allows you—even on the top of a new page—to see whether the line is the first of a new paragraph or the continuation of an old paragraph.

3.3 Typesetting Math as Part of Your Text

Formulas are part of the text, try to use them like nouns and be sure to include punctuation to keep the text flowing. For example, the pythagorean theorem,

$$a^2 + b^2 = c^2, \tag{3.1}$$

is typically taught at school.

Other random notes:

- Note that as in Section 3.2, before and after a formula, empty lines will add space and indentation, and possibly page breaks, or figures. So if a formula is part of a paragraph, do not add empty lines around it in the code.
- Number all equations. It makes it easier to refer to them for others, even if you do not refer to them in your own text.
- While we're at it, do not use the `equation` or `eqnarray` environments, use `align` instead.
- Take care to always put all math in the `math`-environment, for example the variable x . Note that it is typeset differently from the letter `x`.
- Use single-letter names for variables in `math` (`math` is not a programming language!). Function names may have multiple characters, such as $\tanh(x)$. Note that they are not typeset in italics. They usually have their own commands, and you can define your own functions, such as `somefunc(x)`. If you don't do this, the kerning will be broken, since L^AT_EX assumes that you're multiplying the variables s , o , m , e , f , u , n , and c and typesets accordingly, compare *somefunc* (`math` mode, without spaces!) and *somefunc* (italics).

3.4 Referencing Sections, Figures, etc.

Use the `cleverref` package to refer to other places in the document, e.g. Chapter 1. The package allows to specify the reference style globally. Therefore, it helps to refer in a consistent way. Otherwise you may quickly have different ways

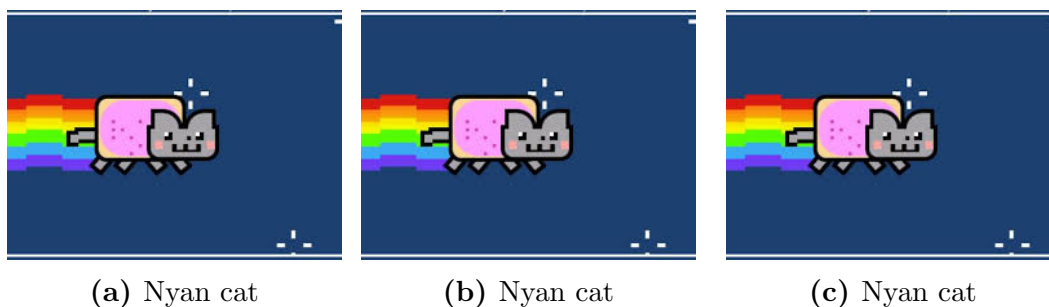


Figure 3.1: Pictures of Nyan cats

of referencing, e.g. *Figure*, *figure*, *Fig.* and *fig.*. The package also ensures that there is no linebreak before the number of the figure/chapter etc.

3.5 Citing Other Works

There are two ways to cite, which depends on whether the authors are an active part of your sentence:

- **muster** showed that...
- This is currently a hot research topic (**muster**)

In \LaTeX , these different citation styles are reflected in the `\citet` and `\citep` commands. **Never use the plain `\cite`!** – it is outdated. This distinction is even more important when you use a numeric style for referencing, where in the second version only the number of the reference is printed. In the first case, a number-only citation does not make much sense.

In your bibliography, you should define common strings, such as conference names, using the `@string` syntax. Otherwise you might quickly end up with a lot of different names for the same conference, which is confusing and inconsistent. (Please have a look at the bibliography file.)

3.6 Floats

3.6.1 Figures

You should allow \LaTeX to place the figures where it wants. The same goes for tables. This is called a `float`, as the figure/table floats around. Look in a well-typeset book, and you'll notice that figures aren't placed randomly in the text, they're rather always on the top of the page if at all possible. \LaTeX has options to

Education		
Major	Duration	Income (€)
CompSci	2	12,75
MST	6	8,20
VWL	14	10,00

Table 3.1: A basic example from the booktabs package.

let figures float to *top* (**t**), *bottom* (**b**), *page of floats* (**p**) and *here* (**h**). The default, **tbp**, is ok for most users, in any case, never use **h**. You can use the `FloatBarrier` command if you want to make sure that a figure is within a specific section.

All figures have a caption from which you can understand most of the figure content and its significance. They are also referenced from the main text (e.g., see Figure 3.1). Do not force linebreaks in captions. Do not deviate from these rules, they are strict.

Figures should provide a short name for every figure in square brackets: []. If you do not define a short name, the whole text of the caption will be displayed in the table of figures, which is usually too long.

Note that Figure 3.1 also shows how to use subfigures.

3.6.2 Tables

Never use vertical lines in tables. See the documentation of the booktabs package for an explanation, or look in your favorite scientific book/magazine to understand that this is a commonplace rule.

A basic example, taken from the booktabs documentation, is given in Table 3.1.

The tabu package The tabu package provides some useful tools for advanced tables.

3.6.3 Algorithms

There are several packages to typeset algorithms. We recommend the `algorithm2e` package. In Algorithm 1 a simple example from the algorithm2e documentation is given.

Algorithm 1: How to write algorithms (Small example from the algorithm2e documentation)

Data: this text
Result: how to write algorithm with L^AT_EX2e initialization;
while *not at end of this document* **do**
 read current;
 if *understand* **then**
 go to next section;
 current section becomes this one;
 else
 go back to the beginning of current section;
 end
end

3.7 Miscellaneous

3.7.1 Enumerations

Short enumerations look ugly in standard L^AT_EX, since they include too much space between the items. A “normal” enumerate list with very short paragraphs:

1. foo bar baz
2. foo bar baz
3. foo bar baz

A list using enumitem to compress the inter item spaces, looks much better for short items:

1. foo bar baz
2. foo bar baz
3. foo bar baz

3.7.2 Front, Main and Back Matters

Titlepage, lists of figures, table of contents page etc. are numbered in roman letters. The introduction should start with page 1 in arabic numerals. The document class `scrbook` does this by default, when `\frontmatter`, `\mainmatter` and `\backmatter` are placed properly.

3.8 Compiling The Document

L^AT_EX is a bit archaic to work with, since with every pass through your document, some intermediate information is written to files, which are then processed by other programs. Keep in mind that everytime a new pass through your document incorporates new information (e.g. from BibTeX), page breaks may change, which requires an additional pass through the document. Thus, you usually have to compile the document in three steps:

1. Compile the document using xelatex: `xelatex thesis`
2. Run biber (or if you do not have it, bibtex) to set up bibliography: `biber thesis`
3. Run xelatex twice more to incorporate bibtex references and to get the type-setting right.

If you know what you're doing you can get away with less. Some programs parse the output of L^AT_EX to determine whether additional passes are necessary. Most GUI programs fall in this category, as do `rubber` and `latexmk`.

3.8.1 Issues With This Template

My references can't be found The line including the `biblatex` package says it should use `biber` as a backend. This is a reimplementaion of BibTeX. If you do not have `biber`, or your IDE does not support it, consider switching the backend to `bibtex`.

Appendix

Note that it the missing chapter number, since it is behind the backmatter command.