



TAMPEREEN TEKNILLINEN YLIOPISTO  
TAMPERE UNIVERSITY OF TECHNOLOGY

# **JAAKKO PASANEN**

## **NLP FOR CUSTOMER SUPPORT AGENT**

Master of Science thesis

Examiner: Prof. Ari Visa  
Examiner and topic approved by the  
Faculty Council of the Faculty of  
xxxx  
on 30th July 2014

## ABSTRACT

**JAAKKO PASANEN:** NLP for Customer Support Agent

Tampere University of Technology

Master of Science thesis, xx pages, x Appendix pages

xxxxxx 201x

Master's Degree Programme in xxx Technology

Major:

Examiner: Prof. Ari Visa

Keywords:

The abstract is a concise 1-page description of the work: what was the problem, what was done, and what are the results. Do not include charts or tables in the abstract.

Put the abstract in the primary language of your thesis first and then the translation (when that is needed).

# TIIVISTELMÄ

**JAAKKO PASANEN:** Luonnollisen kielen ymmärrys asiakapalveluagentilla  
Tampereen teknillinen yliopisto  
Diplomityö, xx sivua, x liitesivua  
xxxkuu 201x  
xxx koulutusohjelma  
Pääaine:  
Tarkastajat: Prof. Ari Visa  
Avainsanat:

The abstract in Finnish. Foreign students do not need this page.

Suomenkieliseen diplomityöhön kirjoitetaan tiivistelmä sekä suomeksi että englanniksi.

Kandidaatintyön tiivistelmä kirjoitetaan ainoastaan kerran, samalla kielellä kuin työ. Kuitenkin myös suomenkielisillä kandidaatintöillä pitää olla englanninkielinen otsikko arkistointia varten.

## PREFACE

This document template conforms to Guide to Writing a Thesis at Tampere University of Technology (2014) and is based on the previous template. The main purpose is to show how the theses are formatted using LaTeX (or  $\text{\LaTeX}$  to be extra fancy) .

The thesis text is written into file `d_tyo.tex`, whereas `tutthesis.cls` contains the formatting instructions. Both files include lots of comments (start with `%`) that should help in using LaTeX. TUT specific formatting is done by additional settings on top of the original `report.cls` class file. This example needs few additional files: TUT logo, example figure, example code, as well as example bibliography and its formatting (`.bst`) An example makefile is provided for those preferring command line. You are encouraged to comment your work and to keep the length of lines moderate, e.g. <80 characters. In Emacs, you can use `Alt-Q` to break long lines in a paragraph and `Tab` to indent commands (e.g. inside figure and table environments). Moreover, tex files are well suited for versioning systems, such as Subversion or Git.

Acknowledgements to those who contributed to the thesis are generally presented in the preface. It is not appropriate to criticize anyone in the preface, even though the preface will not affect your grade. The preface must fit on one page. Add the date, after which you have not made any revisions to the text, at the end of the preface.

Tampere, 11.8.2014

On behalf of the working group, Erno Salminen

# CONTENTS

1. Introduction . . . . .	2
2. Writing style . . . . .	4
2.1 Using LaTeX . . . . .	4
2.2 Text . . . . .	4
2.3 Figures . . . . .	5
2.4 Tables . . . . .	6
2.5 Mathematical notations . . . . .	8
2.6 Programs and algorithms . . . . .	9
3. Referencing styles . . . . .	11
3.1 In-text citations . . . . .	11
3.2 Bibliography . . . . .	12
3.2.1 Header at 3rd level . . . . .	13
3.2.2 Another header at 3rd level . . . . .	13
4. Conclusions . . . . .	14
Bibliography . . . . .	15
APPENDIX A. Something extra . . . . .	16
APPENDIX B. Something completely different . . . . .	17

## LIST OF FIGURES

2.1	Diagrams should be edited before publication. . . . .	6
2.2	Example of subfigures . . . . .	7

## LIST OF TABLES

2.1	Example of evaporation conditions in a thin film structure. . . . .	7
3.1	Necessary bibliographic information. . . . .	12

## LIST OF ABBREVIATIONS AND SYMBOLS

CC license	Creative Commons license
LaTeX	Typesetting system for scientific documentation
SI system	Système international d'unités, International System of Units
TUT	Tampere University of Technology
URL	Uniform Resource Locator

$a$	acceleration
$F$	force
$m$	mass

The abbreviations and symbols used in the thesis are collected into a list in alphabetical order. In addition, they must be explained upon first usage in the text.



# 1. INTRODUCTION

This document template conforms to Guide to Writing a Thesis Tampere University of technology (TUT) *Thesis Writing Guide in English* 2013. A thesis or a report typically include the following chapters:

- Title page
- Abstract
- Preface
- Contents
- List of abbreviations and symbols
- 1. Introduction
- 2. Theoretical background
- 3. Research methodology and materials
- 4. Results and analysis (possibly split into separate chapters)
- 5. Conclusions
- References
- Appendices (if applicable)

Each of these chapters starts from a new page in a thesis. The titles of chapters from 1 to 5 are provided as examples only. You should use more descriptive ones. The title page of your thesis features the University's logo (approximately 8 cm wide), your name, the thesis title and type. The name of the examiner is included at the bottom right corner of the title page. The table of contents lists all the numbered headings after it, but not always the preceding headings.

Introduction outlines the purpose and objectives of the presented research. The background information, utilized methods and source material are presented next at a level that is necessary to understand the rest of the text. Then comes the discussion regarding the achieved results, their significance, error sources, deviations from the expected results, and the reliability of your research. Conclusions is the most important chapter. It does repeat the details already presented, but summarizes and them and analyzes their consequences. List of references enables your reader to find the cited sources.

This document is structured as follows. Chapter 2 discusses briefly the basics of writing and presentation style regarding the text, figures, tables and mathematical notations. Chapters 3 and 4 summarize the referencing basics and the whole

document. There are two example appendices as well (Appendix A and B).

## 2. WRITING STYLE

Effective written communication requires both sound content and clear style. Keep the layout of your thesis neat and pay attention to your writing style.

### 2.1 Using LaTeX

This document serves as an example, rather than tutorial, since there are plenty of those available, see for example Mittelbach et al. 2004; Oetiker et al. 2014; *LaTeX* 2013. The source files are compiled with the command `pdflatex d_tyo.tex` (or `latex d_tyo.tex`). Depending on your bibliography style, you might also need to run `bibtex d_tyo`.

Unfortunately, compilation may fail, if your LaTeX environment does not have all the needed packages installed. If you cannot install them, you can momentarily disable them by putting them inside comments in the file `tutthesis.cls`:

```
% \usepackage{hyperref}
```

Of course, depending on the package, some of the features are also disabled and you must remove those parts from your tex file. Spellchecking can be done, for example, by running `aspell -t -c d_tyo.tex`.

Moreover, certain characters need escape symbols, e.g. per cent should be written as `\%` (otherwise it starts a comment), underscore as `\_`, tilde as `\~{}`, hash as `\#`, dollar sign as `\$`, and opening brace as `\{` and so on. Note that double backslash `\\` makes a line break.

### 2.2 Text

A thesis is written with a single-column layout on one- or two-sided A4 sheets (210 mm x 297 mm). The font type of the body text is usually Times New Roman and the

font size is 12 pt. The spacing is 1.2 and the text is fully justified and hyphenated. You do not have to indent the paragraphs.

Arial 18 pt font is used for the headings in this guide, and there is a 42 pt space above and below. The font size of subheadings is 14. There is an 18 pt space above subheadings and 12 pt space below them.

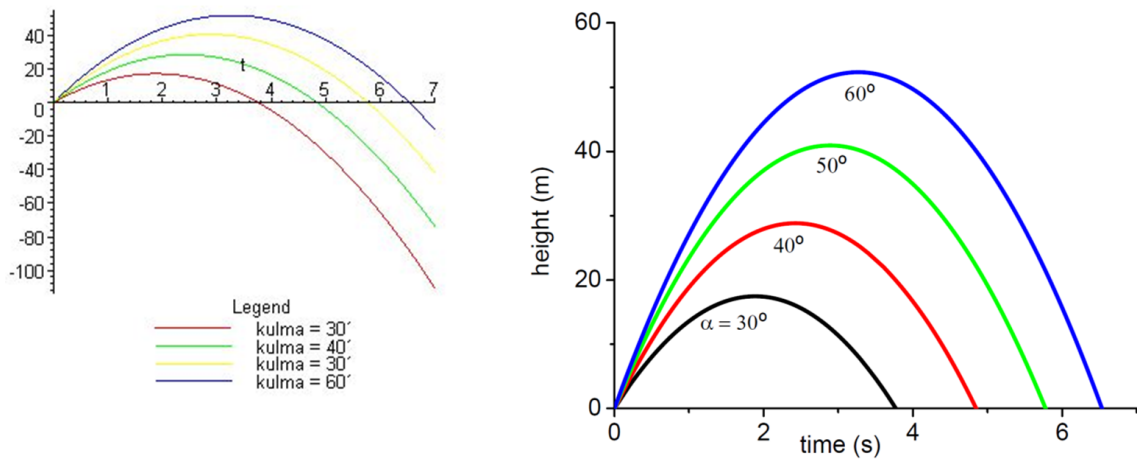
Brief basics of writing style are:

- Always think of your reader when you are writing and proceed logically from general to specific.
- Highlight your key points, for example, by discussing them in separate chapters or presenting them in a table or figure. Use *italics* or **boldface** for emphasis, but don't overdo it. Moreover, this template uses `teletype` font for LaTeX specific names, such as commands and files.
- Avoid long sentences and complicated statements. A full stop is the best way to end a sentence.
- Use active verbs to make a dynamic impression but avoid the first person pro-noun "I", except in your preface.
- Avoid jargon and wordiness. Use established terminology and neutral language.
- The minimum length of chapters and subchapters is two paragraphs, and you need to consider the balance of chapters. Paragraphs must always consist of more than one sentence.
- Do not use more than three levels of headings, such as 4.4.2.
- Do not use too many abbreviations. Use capital and small letter consistently

Sometimes ending a section with list is considered as bad style. Therefore, it is better to have some text after it.

## 2.3 Figures

You must refer to all the figures in the body text. The reference should preferably appear on the same page as the actual figure or before it. Figures and tables must be numbered consistently thesis and primarily placed at the top of the page, but



**Figure 2.1** Diagrams should be edited before publication. The diagram on the right is an edited version of the one on the left.

you are free to decide where they fit best. Never start a chapter with a figure, table or list.

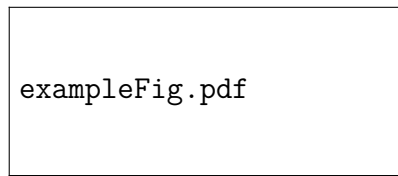
Figures and the caption are either consistently centered (or aligned to the left). The caption is placed under the figure and always on the same page as the figure. All figures must be explained in the body text, so that readers know what they are supposed to notice. Figures generated by analysis software usually need further editing, see Figure 2.1 for example. The figures should be in the same language as other text (even if Figure 2.1 violates this recommendation). The recommended font size is the same as that of the body text but no smaller than 10 pt. The figures must be readable, even if your thesis is printed in greyscale.

The basic command `latex` accepts only encapsulated postscript (eps) format. Therefore it is usually easiest to compile with `pdflatex` which can handle `*.png`, `*.jpg` and `*.pdf` formats. Eps and pdf are recommended since their support vector graphics (zooming). For example Figure 2.1 is in pdf format.

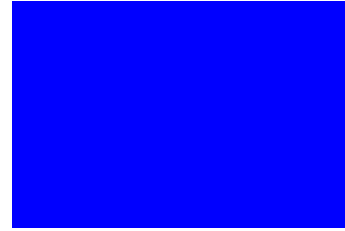
LaTeX has a package `subfigure` to layout multiple figures together, for example Figure 2.2(a) to 2.2(c). There might be newer packages as well, but this is already a quantum leap ahead of some other, unnamed word processors.

## 2.4 Tables

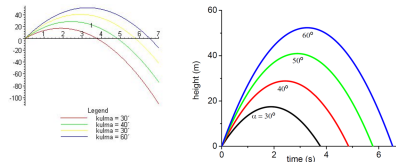
Tables have numbered captions, see Table 2.1 for example. The caption is placed on the same page but above the table, unlike the captions that accompany figures.



(a) Option `draft` helps to make a placeholder box for figure. The file must exist.



(b) Another way to make a placeholder box with the command `rule`.



(c) Actual figure included but scaled down.

**Figure 2.2** Example of subfigures. Pay attention to how nicely they are laid out and their neat subcaptions.

You must refer to all the tables in the body text. In addition, you must discuss the content of any tables in the body text to ensure that readers understand their relevance.

!!!TODO: translate the thin film explanation in English

**Table 2.1** Example of evaporation conditions in a thin film structure.

Substance	Thickness (nm)	Correction coefficient	Pressure (mbar)	Temper- ature (°C)	Current (mA)	Speed (nm/s)
SiO <sub>2</sub>	181.0	1.10	$3.0 \cdot 10^5$	90.6	20-23	0.2
TiO <sub>2</sub>	122.1	1.55	$1.5 \cdot 10^4$	91.1	100-93	0.1

Often it is better to create the table in, e.g. MS Excel, and import it as .eps or .pdf file, for example, when you calculate some of the values automatically.

The titles of the columns in your table are specified and data is added to the table. You can use boldface to highlight the titles and use a double horizontal line to separate the from the rest of the table. The order of the columns and rows must be carefully considered. Do not surround all the cells with a border, as it may make your table harder to read. Put a line on top and bottom of the table. You can add a horizontal line between every 4 – 5 rows, if the data is not grouped into categories. If the table is large, the rows should be numbered if you plan to refer to the rows in

the body text.

The numbers are right aligned (optimally lined up at the decimal point) for easy comparison. You should preferably use SI units, established prefixes and rewrite large numbers so that the power of ten should be placed in the title of the column instead of each row, if possible. More suggestions can be found in Salminen Nov 2009 (updated Aug 2012).

## 2.5 Mathematical notations

Numbers are generally written using numerals for the sake of clarity, for example “6 stages” rather than “six stages”, which is nevertheless strongly preferred to “a couple of stages”. You should also use a thousand separators<sup>1</sup>, i.e. instead of 55700125 write 55 700 125. Never omit the leading zero in decimals. For example, it is correct to write “0.5” and wrong to write “.5”. A comma is used as a decimal separator in the Finnish language and a period in the English language.

Like numbers, it is advisable to abbreviate units of measurement. There is a space between the number and the unit, but you should keep them on the same line. It is better to compile a table or graph than include a great deal of numerical values in the body text. Use precise language and put numbers on a scale (small, fast, expensive).

Use generally known and well defined concepts and standard conventions and symbols for representing them. New concepts should be defined when they appear in the text for the first time. Upper case and lower case letters mean different things in symbols and units of measurement. Do not use the same symbol to mean different things.

Newton’s Second Law can be presented in the following way:

$$ma = F, \tag{2.1}$$

where  $m$  denotes the mass of an object,  $a$  means acceleration, and  $F$  means force. Please note that all the variables must be defined at the point of their first appearance. All sentences end with a punctuation mark, and the main elements of a sentence are separated by a comma in accordance with the rules of English grammar.

---

<sup>1</sup>Use tilde ~ in LaTeX and a special character *non-breaking space* in MS Word

Mathematical formulas are numbered, if they are written on separate lines and referred to in the main body of the text. The number is usually put in parenthesis and right aligned, see equation 2.1 for example. Occasionally mathematical notations are preceded by an identifier, such as 'Definition 1' or 'Theorem 1' Ruohonen 2009. Simple formulas may be displayed within the body of the text without numbering.

Do not start a sentence with a mathematical symbol but add some word, such as the name or type of the symbol, in front of it. Variables, such as  $x$  and  $y$ , are generally presented in italics, whereas elementary functions, special functions and operators are not:

$$\sin(2x + y), \operatorname{grad} T, \operatorname{div} B, \lim (x^2 - 1)/(x + 1).$$

At first, it is better to rely on the automated formatting of an equation editor. You may have to make compromises between logical clarity and readability.

LaTeX is the best editor for writing also the more complex equations, such as

$$G^+(t, t') = \int G^+(E) \exp[-iE(t - t')/\hbar] dE. \quad (2.2)$$

## 2.6 Programs and algorithms

Codes and algorithms are written using monospaced font, such as Courier New, Consolas or their variations. If the length of the code or algorithm is less than 10 lines and you do not refer to it later on in the text, you can present it similarly to formulas. Here's an example showing a snippet from the makefile. These commands were written directly to the .tex file and appear without numbering.

```
all: ${TARGET}.tex
    pdflatex ${TARGET}.tex
    bibtex ${TARGET}
    pdflatex ${TARGET}.tex
```

If the code is longer but shorter than a page, you present like a figure (Program 4.1) titled "Program" or "Algorithm". You should add some comments to the code and indent it consistently. The actions performed by the code must be outlined in broad terms in the body text. Line numbers make it much easier to refer to the code in the text.

LaTeX has a package listings Heinz et al. 2006 which can handle code very conveniently, include real code files, add row numbers, and highlight the reserved words.



Program 2.1 shows another example which is included from a separate file *example\_code.c*, and includes both line numbers and a code numbering.

```
void jarjesta( Kirjainpari taulukko[], int koko )
2 {

4   // Järjestetään taulukko siten, että jokaisella kierroksella
   // valitaan alkio, joka kuuluu taulukossa seuraavaksi ja
6   // siirretään se oikealle paikalleen.
   for( int i = 0; i < koko; ++i )
8       {
           cout << "Hello World";
10          // Etsitään pienin eli lähinnä aakkosten alkua oleva
           // kirjain lopputaulukosta
12          int pienimmanKohta = i;
           for( int j = i; j < koko; ++j )
14              {
                  if( taulukko[ j ].korvattava
16                     < taulukko[ pienimmanKohta ].korvattava )
                      {
18                          pienimmanKohta = j;
                      }
20              }
           // Vaihdetaan pienin alkio omalle paikalleen
22          Kirjainpari tmp          = taulukko[ i ];
           taulukko[ i ]              = taulukko[ pienimmanKohta ];
24          taulukko[ pienimmanKohta ] = tmp;
       }
26   return;
}
```

**Program 2.1** *Example of algorithm. Variable koko is emphasized to highlight some important aspect.*

### 3. REFERENCING STYLES

Different referencing styles determine how you create 1) in-text citations and 2) the bibliography. Two common referencing styles are presented in this chapter:

1. Numeric referencing (Vancouver system), such as [1],[2]...
2. Name-year system (Harvard system), such as (Weber 2001), (Kaunisto 2003)...

A numeric reference is inserted in square brackets [ ], whereas the last name of the author and the year of publication are given in round brackets ().

Both styles are acceptable, but the conventions for referencing vary between disciplines. You must pick one and use it consistently throughout your thesis.

#### 3.1 In-text citations

In-text citations are placed within the body of the text as close to the actual citation as possible. The citation is generally placed within the sentence before the full stop. LaTeX has a command `\cite` for this Oetiker et al. 2014, p. 85. See the tex file for additional remarks for Harvard style citations.

Weber argues that [1].  
 Cattaneo et al. introduce in their study [2] a new...  
 The result is ... [1, p. 23]. One must also note... [1, s. 33-36]

In accordance with the presented theory ... (Weber 2001).  
 It must especially be noted... (Cattaneo et al.).  
 Weber (2001, p. 230) has stated...

Based on literature in the field [1,3,5]...  
 Based on literature in the field [1][3][5]...  
 The topic has been widely studied [6-18]...

...existing literature (Weber 2001; Kaunisto 2003; Cattaneo et al. 2004) has...

**Table 3.1** *Necessary bibliographic information.*

#	Numeric system	#	Name-year system
1.	authors,	1.	authors,
2.	title,	2.	(year in parentheses)
3.	publisher,	3.	title,
4.	year of publication,	4.	publisher,
5.	pages,	5.	pages,
6.	URL, if applicable	6.	URL, if applicable

## 3.2 Bibliography

The entries must include all the details listed in Table 3.1.

Formatting examples of an journal article in bibliography are provided below, first in the numeric style and then the name-year style.

- [100] K. Keutzer, A.R. Newton, J.M. Rabaey, A. Sangiovanni-Vincentelli, System-level design: orthogonalization of concerns and platform-based design, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol.19, no.12, Dec 2000, pp.1523-1543.

Keutzer, K., Newton, A.R., Rabaey, J.M. & Sangiovanni-Vincentelli A. (2000). System-level design: orthogonalization of concerns and platform-based design. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. Vol.19(12), s.1523-1543.

Your references are listed at the end of your thesis in alphabetical order based on the first author's last name. If the author is unknown, alphabetize the source using the corporate author or title.

LaTeX has two ways for making reference list

1. using automated Bibtex tool
2. manually

The tex source of this document has both versions, and the other is in comments. Bibtex<sup>1</sup> formats the reference list according to a setup file, which is usually provided

---

<sup>1</sup><http://ctan.org/pkg/bibtex>

by academic journals. You need to write the basic information into `.bib` file. Citations in your tex file are looked up by bibtex and it produces the list of references automatically.

### **3.2.1 Header at 3rd level**

Some text...

### **3.2.2 Another header at 3rd level**

Section 3.2.1 cannot appear alone, but needs some company (i.e. 3.2.2).

## 4. CONCLUSIONS

This template and the matching Latex document class file together with general writing guidelines should help achieving a consistently formatted and clear documents. Similar template is also available for MS Word.

Every writing and presentation must have a conclusion. This fact is here emphasized by having this short and rather artificial summary also in this template. A concise summary table is a good way for providing an overview of the most important points.

## BIBLIOGRAPHY

- Heinz, C., B. Moses, and J. Hoffmann (2006). *listingsäö - Typeset source code listings using LaTeX*. Comprehensive TeX Archive Network. Available: <http://www.ctan.org/pkg/listings>.
- LaTeX* (2013). 706 p. Available: <http://en.wikibooks.org/wiki/LaTeX/>. Wikibooks.
- Mittelbach, F. et al. (2004). *The Latex Companion*. 2nd. 1120 p. Boston, USA: Addison-Wesley.
- Oetiker, T. et al. (2014). *The Not So Short Introduction to LATEX2 $\epsilon$  - Or LATEX2 $\epsilon$  in 157 minutes*. Version 5.03, 171 p. Available: <http://www.ctan.org/tex-archive/info/lshort/english/>. The Comprehensive TeX Archive Network (CTAN).
- Ruohonen, K. (2009). *Matemaattisen tekstin kirjoittaminen*. 7 p. Available: <http://math.tut.fi/~ruohonen/D-tyo-ohje.pdf>. Tampereen teknillinen yli-opisto.
- Salminen, E. (Nov 2009 (updated Aug 2012)). *Practical advice for writing publications*. 101 p. Available: [http://www.cs.tut.fi/~ege/Misc/salminen\\_figures\\_styles\\_v15.pdf](http://www.cs.tut.fi/~ege/Misc/salminen_figures_styles_v15.pdf). Tampere University of technology.
- Thesis Writing Guide in English* (2013). Available: <https://www.tut.fi/pop> > Study info > Master's thesis > MSc thesis guidelines. Tampere University of Technology.

## **APPENDIX A. SOMETHING EXTRA**

Appendices are purely optional. All appendices must be referred to in the body text

## **APPENDIX B. SOMETHING COMPLETELY DIFFERENT**

You can append to your thesis, for example, lengthy mathematical derivations, an important algorithm in a programming language, input and output listings, an extract of a standard relating to your thesis, a user manual, empirical knowledge produced while preparing the thesis, the results of a survey, lists, pictures, drawings, maps, complex charts (conceptual schema, circuit diagrams, structure charts) and so on.