

How-To write a Bachelor or Master Thesis in the Pattern Recognition Group*

Disclaimer *This document contains information and recommendations exclusively for students writing their BSc or MSc thesis in the Pattern Recognition Group. This document does not apply for students of other groups.*

Formal Requirements and Tools

- **Start:** Once you have agreed with your supervisor on a topic and on a start- and end-date, fill in the corresponding form and complete the respective procedure:
 - For a BSc thesis, the form is available from the *Studienfachberatung*.
 - For a MSc thesis, follow the *JMCS process*.
- **Presentation:** Every thesis must be presented. Usually this occurs during the semester, within the regular seminar organized by the Pattern Recognition Group. You should agree with your supervisor in the initial stage of the thesis on where and when to present your work.
- **Word Processing:** A thesis has to be written with the aid of \LaTeX not with Microsoft Word (or similar word processing software). For BSc theses a sample template is available on the PRG website. For MSc theses, a sample template is available on the JMCS website.
- **Length:** For BSc theses we expect a length of about 30 pages, for MSc theses about 50 pages are appropriate. These are – of course – rough rules of thumb. The actual number of pages of the thesis depends on the type and content of the scientific work.
- **Language:** BSc theses can be written in German or English (English is highly preferred). MSc theses must be written in English.
- **Tools:** We recommend the following tools:
 - **Overleaf**, **TexMaker**, or similar for \LaTeX editing
 - **BibDesk**, **Zotero**, or similar for managing the bibliography
 - **GraphViz** for drawing graphs. Alternatively use **draw.io**, **OmniGraffle**, **LucidChart** (partly fee-based software and partly not available for all platforms)

*Version 1.1, April 2023

Content

A BSc or MSc thesis consists of the following mandatory elements (the content of these elements is briefly outlined below):

1. Cover Page
2. Abstract
3. Table of Contents
4. Major Part (often subdivided into Introduction, Main Part and Conclusion)
5. Appendix
6. Bibliography
7. Declaration of Authorship.

Important: BSc or MSc theses do *not* contain a reflection. Hence, do not describe your erroneous paths or what you would do differently today – a thesis should not read like a travelogue or work log.

A common question is the level of detail the thesis should achieve – or in other words, what level of knowledge one can expect the reader to have. The rule of thumb is that the more important a concept is to your work, the more detailed you should actually describe that concept. For example, if *graphs* are very important and build the basis to your work, you should formally introduce and thoroughly explain the concept of a graph. On the other hand, if you only *use* a certain concept (and this component is virtually interchangeable and not a central part of your own work), you can explain this concept in a few sentences – this is often true for certain algorithms used in the evaluation (e.g., an SVM classifier). But be careful. For example, if you see that certain parameters/details of the concepts used cause large differences in the empirical evaluations, you may need to explain and define these parameters in more detail, even though you are only "using" the concept.

- **Cover Page:** The cover page contains the most important data about the institution, the author and the written work, so that all necessary information can be grasped at a glance (title and subtitle, type of thesis, faculty and university, author's name and origin, as well the names and affiliations of the supervisors). We refer to the templates for details.

The title should inform what the thesis contains. The title can also be provided with a subtitle. Here, the title consists of only a few words and should arouse the reader's interest. The subtitle gives more information about the analysis or topic direction. It should not repeat the information of the title. Furthermore, the title (in normal font) should not be longer than one line.

- **Abstract:** An abstract is intended to give the reader a brief summary and thus provide an overview of the main points of the thesis. The abstract includes the most important points from the introduction, methods, results, substantive new findings, and conclusion. An abstract should be

precisely worded and self-contained. It is limited to approximately 200 words and should not be longer than one page.

One writes the abstract at the very end of the thesis production process.

- **Table of Contents:** The table of contents reflects the structure of the thesis. This gives a first impression of the content, and thus also of the quality of the work. The table of contents makes it easy to check if topics are balanced across the thesis and if the titles of sections and subsections are consistent.

Outline sections include at least two separate sections, i.e., Section 2.1, for instance, has to be followed by at least Section 2.2.

The hierarchy of sections should be limited to a maximum of three levels. Thus, Chapter 2 could be divided, for instance, into Sections 2.1 and 2.2, and Section 2.x into Subsections 2.x.1 and 2.x.2. The final division into subsubsections is then on level 2.x.x. Note, however, that the provided template does not show the numbering on this level anymore and subsubsections are not listed in the table of contents either.

The bibliography and appendix should also be indicated in the table of contents with page numbers.

- **Major Part:** A thesis usually consists of three major parts: an introduction, a main part, and a conclusion. The main part can in turn be subdivided into several chapters.

- **Introduction:** A typical introduction consists of three parts: The context, the goal, and an outline of the thesis.

In the very first part of the introduction the topic should be placed in larger context by embedding the topic in a top down manner.

For example, one can describe that the thesis is in the field of Artificial Intelligence (AI). Then one describes in a few paragraphs what AI encompasses – here one mentions, for instance, that Pattern Recognition (PR) is also one of the sub-branches of AI. Then one goes down a level and describes PR in a few paragraphs as one of the many sub-branches of AI. Next, one goes down a further level and explains that PR can be subdivided into, for example, Classification or Clustering or into Statistical PR or Structural PR. Depending on where the actual topic is located, one finally describes in a few paragraphs this particular sub-branch.

Next, the introduction should contain a clear statement of the research question. That is, one has to introduce the major goal (and possible subgoals) of the project and ultimately answer the question what one wants to achieve with the thesis. This also includes a clear motivation and description of the scientific relevance of the research question.

Finally, at the end of the Introduction one should briefly show how the further structure of the thesis is organized. This includes the flow, what will be dealt with in which chapter.

Before one begins the actual writing of the thesis, it can be helpful to write down in bullet points what should be in the introduction

to have clarity about where you are going. The final version of the introduction, however, is often written after one has finished the main part of the thesis.

Important: In contrast to the abstract, the introduction is not a summary in which the results are anticipated, but a content-related introduction that should invite the reader to read on. What is the topic of the paper, why is it relevant to address it, what is the goal of the thesis, and how is the thesis organized? These are questions that should be answered in the introduction.

As a guideline for the size of the introduction, it should be about 10% of the total thesis length, i.e., for a 30-page thesis, the introduction should be about 3 pages.

- **Main Part:** This part represents the centerpiece of the work. An examination of the topic takes place and the reference to the research question should always be present. The selected literature and the relevant sources serve as evidence and description of the topic.

The structure of the main part depends on the research question, the method and the topic. In our group, many theses are organized as empirical theses that focus on the development and research of a novel method or algorithm applied to some data sources to solve a specific problem.

The main part of the thesis can be – for instance – divided into three major chapters:

- * In the first chapter, the research question is embedded in a theoretical context and the current state of research in the literature is thoroughly presented in its main points. The research question is formally specified and it is shown what science has found out about the topic so far. The theoretical discussion also serves to develop the hypothesis. Which relevant theories and explanations are used to answer the research question? Which hypotheses can be derived from this for the research work?
 - * The second chapter describes in detail the developed solution and the researched method. Typically, this chapter presents and discusses developed algorithms, the implemented framework (including code fragments), as well as data sources and data structures.
 - * In the third chapter, the empirical results are clearly presented and critically discussed. In addition, an attempt can be made to place them in a broader (theoretical) context. What are the results of the empirical investigation? Could the initial hypotheses be confirmed or do they need to be modified? How can they be assessed against the background of existing research findings? What are the strengths or weaknesses of the chosen empirical approach? Which quality criteria are fulfilled by the obtained results?
- **Conclusion:** The conclusion rounds off what has been written and, together with the introduction, forms the framework of the entire thesis. The central questions should be taken up again and discussed.



Figure 1: On <https://dblp.org> one can download BibTeX records of papers.

The developed method is summarized and the main results and other important content should be presented again and one should comment critically on them. Here one shows what answers one has found to the questions and review where shortcomings and problems of the work lie.

Open questions and an outlook to future research should also be formulated. The future work might address the following questions: Is the main question posed in the introduction completely answered in the present thesis? What is the authors own comment? Does the author find the developed solution convincing? Are there any ambiguities? What could be investigated or improved following this or what investigations would have to follow? Where are further findings missing?

Approximately 10% of the total length of the thesis is devoted to the conclusion, i.e., with a total of 30 pages of text, the guideline for the conclusion is about 3 pages.

- **Appendix:** Extra material may be placed in an appendix that appears after the conclusion. Typical content of an appendix is, for instance, part of the code of the developed solution or complete sets of figures that have been produced in the empirical investigation. That is, in the main part of the thesis one shows only the highlights of the code or exemplarily a few figures which support the main findings and then one refers to the corresponding sections in the appendix for the sake of completeness.
- **Bibliography:** One should use BibTeX to define the bibliography. To make the references consistent (e.g. *Pattern Recognition Letters* vs. *Pat. Recognit. Lett.* vs. *PRL*), we strongly recommend to choose a platform (and stick to it) from which one can find and load the vast majority of the references actually cited in the thesis as BibTeX records (we recommend to use <https://dblp.org> – see Fig. 1). References that one does not find on the chosen platform, should be carefully defined such that they match the style and level of detail of the other references.

As a very rough guideline for the number of references in the bibliography, the page length of the thesis can be considered, i.e., if the topic is discussed on 25 pages, the bibliography should consist of approximately 25 references.

- **Declaration of Authorship:** The thesis must also contain the signed declaration of authorship (“Erklärung”), in which the candidate confirms that she/he personally authored the work. Print, sign, scan, and include the declaration at the very end of the thesis.

Scientific Writing

- **Rules:** One needs to know the grammar rules, even the “small” rules about when to capitalize and where to put commas. When you deviate from expected rules, you draw a reader’s attention and a reader’s attention is a precious gift, not to be wasted. In scientific writing, you want your ideas to receive the attention, not your writing. We suggest Strunk and White *The Elements of Style* as a starting point for reading.
- **Tense:** In general, a scientific thesis is written in the present tense. For example: *With the help of this experiment we show . . .* or *Ramon et al. prove in their seminal paper [17] that . . .*
- **Form of language:** There are various forms of language. For scientific texts, the standard high-level language or technical language is the right choice. Colloquialisms, on the other hand, should be avoided. An objective attitude also characterizes a good scientific investigation.

To write objectively, avoid

- subjective exaggerations (e.g., *The results show how perfectly the new system works.*)
 - meaning-enhancing and subjective terms (e.g., *The many tests naturally led to very different results.*)
 - casual evaluations (e.g., *Unfortunately, this approach proved to be completely pointless.*)
 - the expression of personal enthusiasm and own opinions (e.g., *This statement can be wholeheartedly agreed with from the author’s point of view.*)
 - colloquialism and platitudes (e.g., *So it came as it had to come – the algorithm didn’t work.*)
 - filler vocabulary (e.g., *Although there apparently are proponents of this view, it is not accurate, so to speak.*)
- **Simple words and simple sentences:** Strive for clarity, conciseness, and stringency. Simple words and simple sentences are particularly effective. Avoid long, convoluted, perhaps even warped, phrases that twist, turn and distract from the point one might wish, in this ever-so literary of worlds, to make in the most elegant way possible. Do not use long sentences with complex grammar and unusual words just to show that you can. Particularly in technical writing, simple sentence structure and simple words enhances clarity. There is also rarely a reason to use synonyms for the same word – choose a terminology and stick to it. This reduces the mental load on the reader.
 - **Bullet lists:** Bullet lists can be effective for drawing a reader’s attention to a set of important statements. However, they are not an excuse for writing abbreviated or sloppy prose. Bullet lists should be punctuated consistently. You should use a consistent sentence or phrase structure in each item.

Bad Example:

- First we review the list.
- Define the list
- The list needs to be made.
- List design

Better:

- Review the List
- Define the list
- Make the list
- Design the List

- **Consistency:** In scientific writing you should pay attention to consistency. Never switch between different notations or conventions within your thesis. For some words there are several possible spellings. Once you have decided on a spelling, you should keep it throughout your text and not switch between different options (e.g. *dataset* vs. *data set* vs. *data-set*).

- **Abbreviation:** You should introduce an abbreviation that is not commonly used correctly the first time you use it (e.g., *An electrocardiogram (ECG) was obtained. The ECG was unremarkable.*). After that, you should use it consistently and not switch between the written version and the abbreviation. Exception: In order not to disturb the reading flow, abbreviations at the beginning of a sentence are always written out.

Avoid using different abbreviations for the same word. It is best to first check if there is already a given abbreviation for the word in question. This is better than making up your own abbreviation.

The abbreviation e.g. means “for example” while i.e. means “that is”. Do not mix them up. Both abbreviations are commonly followed by a comma. That is, you should punctuate them as you would the equivalent English phrase.

Note that L^AT_EX puts more white space after a “.” This makes sense in the case of a period. Yet, in case “.” is not a period (for example, a “.” used in an abbreviation) a normal space should be displayed. We recommend to use the tilde symbol “~” rather than a normal space in such cases.

- **Citation:** A citation is an annotation for a sentence. It is not part of the sentence and should play no grammatical role in the sentence. In other words, if you remove the citation, the sentence should still be grammatically correct and complete (Bad example: *As it is defined in [17] ...*).

I recommend avoiding lists of citations like this: *Many researchers have studied these normal forms [17-22]*. It should be clear to the reader why you are citing a work and what connection it has to your work.

In L^AT_EX the tilde symbol “~” has the same semantics as a space (“ ”) except it prevents a line break. So we recommend using tilde (instead of a space) before a citation or reference (in order to prevent a line break directly before a reference or citation).

- **Figures and Tables:** It is important that you reference and describe all tables and figure in your body text (use the \LaTeX commands `ref` and `label`). For example, give a reading example, draw the reader’s attention to particularly interesting results, derive a conclusion from the data or figure shown, etc.

Make sure that the tables are easy to read and neatly formatted (\LaTeX offers countless possibilities for this – consult, for instance, the wikibook *LaTeX/Tables* for more information).

With scans or screenshots of images, the question of copyright is sensitive. It is thus best to use figures and images that you have created yourself. Make sure that they are displayed at a high resolution.

- **Algorithms:** Display algorithms and procedures by means of the \LaTeX packages `algorithm` and `algorithmic` (see template for an example). The displayed algorithms must be described in the body text of the work. That is, thoroughly describe what happens on which line of the algorithm and for what purpose.
- **Code:** Display code (e.g., Python or Java code) by means of the \LaTeX package `minted` (see template for an example). Do *not* screenshot your code fragments.

A common question is how much of one’s code should be shown in the main part of the thesis. As a rule of thumb: As little as possible but as much as necessary to estimate the programming effort. We recommend showing and commenting on key passages or particularly exciting parts of the code only. More complete excerpts from the implemented code can be shown in the appendix.

Sections and Paragraphs

- **Plan Structure first:** A good division of the chapters into sections and paragraphs is crucial for the readability of the thesis.

Before starting the actual writing, the structuring should be discussed with the supervisor – a skeleton of the thesis, showing the division into sections and subsections, and also indicating for each section the core sentences (see below) that will be covered in that section by means of paragraphs, serves as the basis for this discussion.

- **Rules for Sections:** It is quite common to briefly outline the sections at the beginning of each chapter in a preamble and explain what is discussed in that chapter.

Each section and subsection discusses a specific topic and needs a descriptive name (section and subsection titles should fit on a single line). The use of informative titles help readers to know exactly what follows and where to look up specific information.

One uses sections when there are more than two subtopics in a chapter which are best explained independently. Vice versa, one does not use sections and subsections if there are no subtopics in the text, e.g., if the text is very short or the subtopics are strongly related with each other or one does not want to interrupt the reading flow.

- **Rules for Paragraphs:** Within a paragraph, one explains a specific core idea, thought or consideration that belongs to an overarching topic (of the corresponding chapter or section). A paragraph begins on a new line and always consists of more than one sentence.

One can only write a good paragraph if one knows the core idea of this paragraph. One needs to know what one wants to accomplish with the paragraph.

- Give an example?
- Make a clear point and elaborate on it?
- Explain a cause or consequence?
- Introduce an algorithm?
- Describe advantages and/or drawbacks of a method?
- Discuss a table or a figure?

Once one knows what to accomplish with the paragraph, e.g., describe the positive effect of a certain procedure, one can write down 1 to 3 core sentences of the paragraph (this can also be accomplished with shorthand notes).

Successive paragraphs are connected in the best case by linking words. For example, assuming the previous paragraph discusses the time complexity of a given algorithm, then the next paragraph could start, for instance, as follows:

As we have now an algorithm for the Lorem Ipsum problem with $\mathcal{O}(n^2)$ time complexity, the next question is how to ...

Completion

- **Submission:** The final version of a thesis consists of a single PDF file. Printed copies are not needed.
- **Grading:** The assessment of a BSc thesis consists of three grades:
 - *Bachelorarbeit:* Assessment of the final version of the thesis
 - *Seminar:* Assessment of the presentation given in the seminar (see documentation in the seminar for details of the grading scheme)
 - *Anleitung zu wissenschaftlichen Arbeiten:* Assessment of the scientific approach and procedure. In the vast majority of cases, this grade corresponds to the same grade as the grade for the thesis. In exceptional cases, however, the grade may differ – for example, to recognize a high level of commitment, a special innovation that goes beyond the initially defined objectives, or a particularly extensive and well-defined experimental evaluation.

The assessment of a MSc thesis consists of one grade and written statements on the following items:

- Has the candidate reviewed, integrated and cited relevant scientific literatures sufficiently?
- Qualification of the scientific research work of the candidate in terms of work organisation, personal effort, aptitude, creativity.
- Relevance of addressed questions and quality of obtained results. Have the aims of the study been met?
- Analytical abilities and interpretation of the results.
- Editorial formulation and layout of the thesis including figures and tables.
- Independent workability of the candidate.

Normally, we grade the thesis within two to three weeks and offer – if desired – a discussion of the evaluation.

- **Failing and Reject:** If a thesis has serious deficiencies, the thesis will be graded with a grade smaller than 4.0 (insufficient).

Yet, if a thesis has deficiencies that can be corrected within a reasonable period of time, we might reject the initial submission. If the student decides to correct these deficiencies within a defined period of time, a passing grade is still possible (typically a grade of 4.0).