Guidelines for Each Thesis Section

Requirements

- The thesis should have the ACM format (double column).
- We offer an Overleaf template to work with.
- The maximum length is 10 pages (references should start no further than the second column of the tenth page, excluding appendix).
- References follow the ACM format.

1. Abstract

- The function of an abstract is to allow readers to judge whether the paper is relevant to them. It should be 50 to 200 words. Avoid citations.
- The abstract should be clear and concise, giving an accurate view of what has been found, how it relates to former research and why it amounts to state-of-the-art research.
- The five-element organization is the preferred structure of the abstract:
    - A general statement introducing the research area of the topic being investigated
    - An explanation of the specific problem addressed (research gap)
    - A review of existing models and theory (if there is any) in relation to the problem being investigated
    - An outline of the proposed model or approach
    - A summary of whether hypotheses/expectations could be confirmed (**evaluation**) and how this contributes to scientific knowledge
- There should be keywords under the abstract that summarize where the thesis is about.

2. Introduction

- The introduction starts with positioning the topic of the research in a scientific field. This can be done in a task-oriented manner. Examples are: Natural Language Processing, Image Classification and Information Retrieval. It is also possible to categorize the scientific field in a subject-oriented manner, like Medical Science or Agricultural Sciences.
- The introduction continues with a problem statement. This is the problem your research sets out to solve. It provides a context of the value of conducting research in the area. Why is the problem scientifically relevant? Make trade-offs explicit that play a key role.
- There will follow a related work section after the introduction. In the introduction, you will need to mention key paper(s) for your research and how they are related to the research gap. Where is existing knowledge thin? The gap needs to be formulated explicitly.
- Following the research gap, you will formulate a **research question** for your master thesis.
    - Why do you need a research question?
        - The function is to structure your thesis, your work, research and everything you do during the thesis period. It is also meant to make sure your reader (examiner) **always knows** why she is reading something. The main research question with sub-questions are the backbone, or even the whole skeleton, of your thesis. They provide the basis for a smooth reading experience, a clearly built structure and a

convincing argument.

- A tree-like structure of (sub) research questions can work well:
  - The standard setup is: depth 3, branching factor 3, numbered RQs.
  - So: RQ, SRQ1, SRQ2, SRQ3, SRQ1.1, SRQ1.2, ...
  - The **root** is a general question making clear what you will do. Make sure that it is not a yes/no question. It can be helpful to start with "to what extent" and to make sure that it is comparative with respect to a baseline.
  - Your sub RQs provide the **evidence** with which you will answer the root RQ in your conclusion. Make sure that it is a question that is answered by what you did.
  - Your leaf sub RQs are super concrete. Readers should be able to predict the *shape of the answer* (e.g., a PR curve, a set of F1 scores for different systems plus indication if differences with a baseline are significant, etc.).
  - All RQs should be easy to read and understand. The higher in the tree, the more they are understandable by non-experts.
- For every paragraph, every subsection, every section in your thesis, it should be clear to which (S)RQ it is connected. Everything you do in the three months need to be connected to a (S)RQ. And *vice versa*, you need to answer all your (S) RQs.
- What does not count as a research question?
  - All questions for which others have provided the answer already. All things you can find in the literature can thus never be the answer to a RQ.
  - To find the answers to *your RQs*, **you** must do *research*! This means that you work with data, do experiments, code, think, play, experiment.
  - Avoid formulating questions that can be answered trivially, like "*Is it possible*...." (yes, everything is possible, well almost everything).
- You will find two examples of tricky cases below:
  - *What is the optimal...* That is a pretty hard question to answer if it is not constrained very heavily.
  - *Which AI/ML/... method for my XYZ problem leads to higher profits for my company?* Can you actually test that? Which test design is needed for it? Will your organization/company allow you to do that? Maybe it was the *motivation* to start the project for your company, but a motivation is usually not an RQ. Make sure that the question is scientifically relevant.
- For Data Science projects, it is helpful to combine the Explorative Data Analysis (EDA) with your RQs and ask yourself the following questions:
  - Do I have the data, the features and the gold standard train and test sets to answer each of my RQs and SRQs?
  - Is it **crystal clear** what data I need for each of my SSRQs? What does the quality and the "richness/content" of the data need to be?

3. Related Work

- The related work section is not the same as a background section.
  - Sometimes you want to **explain** techniques you use, referring to the literature. This can go into a background section. Such a section is not necessary and should only be used

when it has benefit for an informed audience. When used, a background section follows the related work section.
    - In a related work section, you describe how other people have dealt with similar problems that you are dealing with. What were their techniques? What were their results? What observations did they make that you have used?
- The role of the related work section. Why?
    - Your work is **grounded** in earlier work done by other researchers. That means you use techniques and methods, ways of measuring and evaluating, datasets and terminology used by others working on the same or a related problem before you. All four aspects need to be addressed in the related work section.
    - You want to **compare** your results with those of others. This is especially true for key papers with respect to your research – the *State Of The Art* (SOTA). Those papers need to be referred to in the related work section. It could also lead to a hypothesis, which is the answer that you expect based on previous literature.
    - The previous two points mentioned cover the main function of the related work section. Additionally, you **inform** the reader on the SOTA, the state of the problem and the solutions. You also **show** that you know what you are talking about, that you know what is "for sale" and that you use the best there is.
- The preferred form of the related work section is as follows.
    - Organize your related work section into 3 to 4 subsections.
    - It should be immediately clear to which subsection which (group of) research questions belongs. There should also be a thematic overview in which the papers are clustered.
    - The starting subsection simply lists and describes the *state of the art* regarding your main research question. This will cover roughly 5 to 10 references, where each reference is described in 2 to 3 sentences.
    - The other subsections are usually more technical (and connected to a theme). You use them to describe and refer to technologies and methodologies you will use for answering your research questions.
- You will find some tips on related literature listed below:
    - Use this schema for your citation keys: nnnn:aaaayy. The nnnn refers to the first 4 characters of the first author, the aaaa to the first four characters of the first non stop word of the title and yy to the last two digits of the publication year.
    - Save all pdfs in a literature folder. Use the same name for the file as the bibtex key.
    - Use markers and notes to annotate the pdfs so you can quickly find the important parts back. You can also use Mendeley, Zotero or Docear for this purpose.
    - Write for each article a small synopsis with the 3-4 main points, a short abstract in your own words and 2-3 lines why this is relevant to your research.
- You can use the following questions as a checklist for the related work section.
    - Is there a clear overview of the approaches to the problem and the "state" of the solution?
    - Is there a clear view of the SOTA?
    - Can all things you have read be related to specific research questions?
    - Gives the related work section a convincing image that the author knows the field, that cutting edge solutions and technology are used and that the author places herself in a

tradition to which she adds a little?
    - Is it possible to give a brief exposé on research questions and the state of the art regarding this topic based on what you have read in the related work?

4a. Methodology

- *For Information Systems students, it is not always necessary to have a separate methodology section. You can integrate the approach and results in one section, which allows for more efficient usage of the 10 pages. It depends on the kind of research what is best fitting.*

- You have mentioned how others have used the methods you will use in your related work section. In the methodology, you will go into more details about them.
- This is not a textbook section (readers can find that elsewhere, probably better), nor a place to copy/paste difficult intimidating formulas.
- One function of the methodology is to **describe** what you do. It is not just about what you are using, but how in particular you are using it. The other function is to **justify** what you do. Why did you choose the current methodology over other plausible alternatives? Does the research design probably measure the effect and is it credible in terms of reliability and validity? How did you go about solving or making progress on the research problem? Did you use simulation or analytic models? What important variables did you control, ignore, or measure?
- In particular, you pay attention to that new little thing, that change, that great idea, that you **add** to the existing method.
- A good paper will make clear the type of research design is used, possible by reference to earlier, similar studies.
- This section can be quite brief. Use most space for your addition. If there is none, the section can be rather brief.

4b. Methodology/Experimental Setup

- *The experimental setup can be either an extensive subsection in the methodology, or it can be made into a separate section. The first approach is traditional but the other is viable as well.*
- Function 1: The user can **replicate** your experiments.
- Function 2: The user gets a (very) **good idea of your used dataset(s)**.
- In the *description of the data* subsection you paste your most insightful graphs from your EDA, next to the basic statistics on your dataset and descriptions (statistical and/or a population density diagram) of your variables.
- In the experimental setup (sub)section you give all the *settings* used in your experiments. All that is needed so that someone with your dataset and your software can replicate your work and obtain more or less (often there are random effects) the same results. Think of:
    - Data pre-processing steps used
    - Hyperparameter settings (for all of them)
    - How you created a train-validate-test split or otherwise did your training and testing
    - Exactly how the used metrics are calculated (think of the difference between micro and macro F1 for classification and how often this is not explicitly stated)

- This (sub)section can be quite long. If you want or need to tell more, consider using the Appendix or a reference to a nicely structured notebook in which all experiments are done.
- *Sometimes, especially if you have quite different, experiments/research questions, it makes sense to interleave the experimental setup and the result sections, so the reader does not get lost. It is then helpful to structure clearly in (sub)subsections.*

5. Results

- The function of the results section is that you give, for each research question, the **outcomes** of the experiment corresponding to a research question in the form of a table or graphic. *In the case of interview-based research, you will show examples and counterexamples by means of quotes.*
- Structure your section in a way that the reader should only read these two things (and can safely skip all else): the question and the table/figure/quote and the (elaborate) caption.
- It is important that tables/figures are unambivalent. There is a need for a perfect caption, perfect labels and smart design of table or figure. The reader should be able to use 100% of her brain to **understand** the outcomes, not to try to figure out what was meant.
- This section can be quite brief (in words). You will answer your research question in the discussion and conclusion.
- Specifically, most good computer science papers conclude that something is so many percent faster, cheaper, smaller, significant or otherwise better than something else. Avoid putting the result there in numbers, handwaving results such as "very" or "small". If you must be vague, you are only given license to do so when you can talk about orders of magnitude improvement. There is a tension in that you should not provide numbers that can be easily misinterpreted, but you do not have room for all the caveats either.

6. Discussion

- The function of the discussion section is to **compare** your results with the *state of the art* and to **reflect upon** the results and the limitations of the study.
- Do not forget to use sub-sections.
- You can start with a sub-section in which you compare the results of your research with previous studies. Be sure to make use of concrete results. Make use of references. You would like to know whether the found effect size is comparable to previous related studies. Be sure to make use of confidence interval or significance, if applicable, to argue that your results are not just random.
- If you did not find what was expected, it is important that you go deeper into the possible reasons for this. Are there comparable studies in which no results was found? What would be the most likely scenarios for not finding the results? It is important that you are able to justify that fixable problems with the research set-up are the most likely reason for not finding a result. It will help to find a theory for your results if you use a second dataset with known results to compare with. In that way, you know whether your results are as expected based on the known dataset.
- The limitations of the study should be noted. Do not make a shortcoming into an excuse. Make it into a strategy for further research. Limitations should be reflected upon by using key concepts like reproducibility, scalability, generalizability, reliability and validity. Make sure that you make clear how you take the limitations into account in the conclusion, if the issues are serious. You can already hint in the discussion at future work to which you come back in the conclusion section.

- Consider whether there are alternative conclusions consistent with the results presented. What is **the value of your research** in light of previous research? Does it indeed fill the research gap? This gives a bridge to the conclusion.

7. Conclusions and Future Work

- This section should not be longer than half a page. You will need to be very sharp and concise.
- The function of the conclusion section is that you **answer** all your research question(s).
- You start with a recap of the scientific relevance and problem statement.
- You then present the answer to the research questions. You will need to (indirectly) rephrase the questions. For example: "*This research aims to answer…*"
- There should be a statement on how the limitations of the study **qualify** the conclusion. If the improvement only shows in specific circumstances, it amounts to a qualification. Make sure that your qualification is properly connected to your statements in the discussion.
- You then go to a statement in which you state what the value of your research has been when compared to the *state of the art*. What does your research add? What are the implications of your answer?
    - Is it going to change the world (unlikely), be a significant "win", be a nice hack or simply serve as a road sign indicating that this path is a waste of time? Are your results general, potentially generalizable or specific to a particular case?
- You finish the conclusion section with future work. You can make a link here to the limitations of the study in the discussion section. Be sure that you argue what is the most promising way forward instead of stating which research "must" be done.