

Universität Heidelberg
Institut für Informatik
Sommersemester 2018
Seminar Embeddings for Data Analysis
Dozent: Dr. Erich Schubert

Seminararbeit
Template for a Seminar Paper

Name: Erika Mustermann
Matrikelnummer: 1234567
Studiengang: Informatik (7. Fachsemester)
Email: erika.mustermann@stud.uni-heidelberg.de
Datum der Abgabe: April 12, 2018

Hiermit versichere ich, **Erika Mustermann**, dass ich die Seminararbeit mit dem Titel **Template for a Seminar Paper** im **Seminar Embeddings for Data Analysis** im **Sommersemester 2018** bei Dr. Erich Schubert selbstständig und nur mit den in der Arbeit angegebenen Hilfsmitteln verfasst habe. Zitate sowie der Gebrauch fremder Quellen, Abbildungen, Texte und Hilfsmittel habe ich nach den Regeln guter wissenschaftlicher Praxis eindeutig als solche gekennzeichnet.

Mir ist bewusst, dass ich fremde Texte und Textpassagen nicht als meine eigenen ausgeben darf und dass ein Verstoß gegen diese Grundregel des wissenschaftlichen Arbeitens als Täuschungs- und Betrugsversuch gilt, der entsprechende Konsequenzen nach sich zieht. Diese bestehen in der Bewertung der Prüfungsleistung mit "nicht ausreichend" (5,0) sowie ggf. weiteren Maßnahmen.

Außerdem bestätige ich, dass diese Arbeit in gleicher oder ähnlicher Form noch bei keiner anderen Prüfung vorgelegt wurde.

Heidelberg, den 12. April 2018

Abstract

Provide a short summary of your presentation and report, for the seminar announcement.

Contents

1. Introduction	1
2. Prior Work	1
3. [Name this section appropriately for your paper]	1
4. Discussion	1
A. Template Introduction	iv
A.1. Language: English or German	iv
A.2. Metadata	v
A.3. Basic structure	vi
A.4. Anti-Plagiarism Declaration	vii
B. Mathematics, Figures, Tables, and Algorithms	viii
B.1. Mathematical Expressions	viii
B.2. Figures	ix
B.3. Tables	x
B.4. Algorithms	x
B.5. Theorems and Proofs	xi
B.6. Utility Commands	xi
C. Tips & Tricks for \LaTeX	xi
D. References and Literature	xii
D.1. Citing Literature	xii
D.2. Citing Web Pages	xiii
D.3. Citing Images	xiii

1. Introduction

This seminar paper discusses the following article:

Andreas Spitz and Michael Gertz. “Terms over LOAD: Leveraging Named Entities for Cross-Document Extraction and Summarization of Events.” In: *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016*. 2016, pp. 503–512. DOI: 10.1145/2911451.2911529

Your task is to summarize and contextualize the work.

Do not just rephrase the work – it is not your task to pretend that you wrote it.

Instead, the assignment is to contribute a *new* view on the existing article, with additional insights on its strengths, weaknesses, and impact. Do not blindly trust every claim in a paper. Every algorithm claims it is better than all others – this cannot be true every time.

It is probably a good idea to first explain the setting:

- What problem do we need to solve?
- What data do we have?
- What are the difficulties?

The following is not a mandatory structure, but may be helpful.

2. Prior Work

Discuss the prior work necessary to understanding the paper *briefly*.

Show that you have properly *researched* related work.

You may assume the readers know basic computer science and the mandatory readings.

3. [Name this section appropriately for your paper]

Describe and explain the main contributions of the paper (you may use more than one section).

4. Discussion

Discuss the impact of the paper for science, its drawbacks, and its strengths.

A. Template Introduction

This is a basic template for a seminar paper or practical report to help you with fulfilling all formal requirements. But it is not, by any means, authoritative, or binding. If you notice an error with the template, you should rather fix it, than leave it as is “because it was already broken in the template”. Instead, we would appreciate improvements.

This template assumes you already used \LaTeX before. It does not explain how to compile a document or such fundamentals. If you have not used \LaTeX before, you probably will need to read through some online tutorials, or a book from the library.

This template consists of four files:

File name:	Contents:
<code>seminar.tex</code>	\LaTeX source
<code>seminar.pdf</code>	\LaTeX output file (this file)
<code>references.bib</code>	Bibliography database
<code>dbstmpl.sty</code>	\LaTeX style file (you do not need to edit this)

A.1. Language: English or German

We encourage you to write in English, but you are allowed to submit your thesis in German. Unfortunately, if you write in German, you will end up using a lot of technical expressions in English, or with some clumsy and uncommon translations. Usually, it will be more elegant to write English instead when working mostly with English sources, and it will be a good opportunity to improve your English writing skills.

To use German as default language, add the option `german` to the style file:

```
\usepackage[german]{dbstmpl}
```

and it will automatically use German for built-in strings such as “Inhaltsverzeichnis”.

A.2. Metadata

First, you need to choose the type and title by uncommenting the appropriate line, and list the instructors / advisors.

```
%\beginnerpractical{Report Title}
%\seminar{Seminar title}{Title}
%\advancedpractical{Report Title}
\date{\today}% Date of submission.
\term{Summer}{2018}% Term and year

\instructor{Prof. Dr. Michael Gertz}
\instructor{Dr. Johanna Geiß}
\instructor{Dr. Erich Schubert}
\instructor{Sebastian Lackner}
\instructor{Andreas Spitz}
```

Last but not least, your personal information:

```
\author{Erika Mustermann}
\matrikelnr{1234567}
\studyprogram{Informatik}{7}
\email{erika.mustermann@informatik.uni-heidelberg.de}
```

These commands will be used for generating the title page and anti-plagiarism declaration.

If your page headline is too long, you can use `\shortauthor{E. Mustermann}` to shorten your name, and `\shorttitle` to provide an alternate, shorter version of the title.

A.3. Basic structure

The document consists of three building blocks, known as the “front matter”, the “main matter”, and the “back matter”.

Front matter will contain the title page and the table of contents:

```
\begin{document}
\frontmatter
\maketitle           % Generates the title page
\eigenstaendigkeitserklaerung % Anti-plagiarism declaration
\tableofcontents    % Table of contents
```

The generated title page **does not have a logo**, because **using the seal is restricted**.

The anti-plagiarism declaration (Eigenständigkeitserklärung), although in German, is something you will have to read, understand, and sign in the printed version. For more details, see Section A.4

The main matter will contain your main contents:

```
\mainmatter
%% Introduction, related work, theory, experiments, conclusions, ...
```

The back matter will contain optional appendixes, and the generated bibliography:

```
\backmatter
%% Optional appendix sections:
%\appendix
%\section{An appendix section}
%% Mandatory bibliography:
\cleardoublepage
\printbibliography % Bibliography
\cdpage           % A page to attach your CD (theses only)
\end{document}
```

A.4. Anti-Plagiarism Declaration

You have to sign the declaration that you adhere to “good academic practice”.

In particular, you should be aware of the university guidelines on academic misconduct, partially reproduced here:¹

§1 Academic misconduct – definition Academic misconduct applies in particular when deliberately, or with gross negligence, false statements are made in scientific and other scholarly studies, the intellectual property of others is violated or their research activity is intentionally harmed in some other way. Misconduct occurs, in particular, in the following cases:

- a) False statements
 - Inventing data
 - Faking data, e.g. by
 - selecting and rejecting undesired results, without making this known
 - manipulating a drawing or figure
 - giving incorrect information in an application for a post or for funding (including false statements about publications in the printing process regarding the journal or other organ);
- b) Violating intellectual property
 - in connection with work protected by copyright, or of essential scientific findings, hypotheses, teachings or research approaches stemming from others:
 - unauthorised use, claiming authorship (plagiarism)
 - exploiting research approaches and ideas, particularly as referee (stealing ideas)
 - appropriating or, without grounds, assuming academic authorship or co-authorship
 - faking content
 - deliberately delaying publication of a scientific or other scholarly study, particularly as editor or referee, or
 - going ahead with the unauthorised publication and enabling accessibility to third parties before the work, the findings, the hypothesis, the teaching or the research approach have been made available to the public;
- c) the claiming of (co)authorship of another person without his/her consent;
- d) the sabotage of research activities (including the damaging, destroying or manipulating of experimental designs, equipment, documents, hardware, software, chemicals, cell and microorganism cultures or other items required by another person to carry out an experiment);
- e) misappropriation of budgetary funds and private grants;
- f) disposal of original data, if this is against statutory regulations or – depending on the discipline – recognised principles of scientific or other scholarly work.

¹For further information, see in English / in German:

https://www.uni-heidelberg.de/university/profile/academic_practice/

https://www.uni-heidelberg.de/universitaet/profil/wissenschaftliche_praxis/

B. Mathematics, Figures, Tables, and Algorithms

\LaTeX largely automates enumeration, referencing, and placement of floating objects such as figures. Because of page layout rules this can sometimes behave a bit surprising, and we recommend that you do not waste too much time to fight these rules, but let \LaTeX perform the layouting. Some important rules to follow:

- Preferably place the float just after or just before the paragraph first referencing it.
- Avoid “here” placement of figures. At the top of a page is fine.
- A `\clearpage` before a new `\section` will keep figures in their section.
- *Always* provide a `\caption` and a `\label`.
- Use `\ref` to reference enumerated items (e.g., `Figure~\ref{fig:abc}`).
- Prefixes like `fig:`, `tab:`, `sec:` for labels reduce duplicate names.

We also provide macros: `\refsec{sec:some-section}`, `\reffig{fig:some-figure}`, `\reftab{tab:some-table}`, `\refalg{alg:some-algorithm}`, ...

B.1. Mathematical Expressions

\LaTeX is known for its high-quality typesetting of mathematical expressions.

They can be used inline (e.g. $x = 42$), or as display math:

$$\int_x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) dx = 1$$

For longer derivations, it is beneficial to use the `align` environment, which allows aligning equations by inserting the `&` symbol before the `=`:

$$\sum_{x=1}^n x = 1 + 2 + 3 + 4 + \dots + n$$

can be rearranged according to Gauss into:

$$\begin{aligned} &= 1 + n + 2 + (n-1) + 3 + (n-2) + \dots \\ &= (n+1) + (n+1) + (n+1) + \dots \\ &= (n+1) \cdot \frac{n}{2} = \frac{n^2+n}{2} \end{aligned} \tag{1}$$

The `align*` environment will not enumerate equations, whereas `\notag` only omits a single number. Use `\label` to name an equation. You can reference the equation using `Equation~\eqref{label}` or the shorthand `\refeqn{label}`.

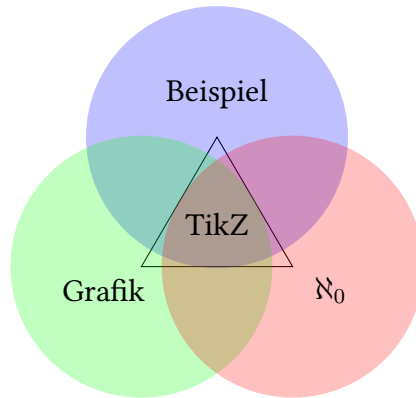


Figure 1: Every figure should have a caption and a label.
Unless you created the figure yourself, always give a reference in the caption.

For example, we get the sentence “According to Equation (1), we can compute the sum of the first n integers in $O(1)$ time.” by writing

According to `\refeqn{eqn:gauss-sum}`, we can compute the sum of the first n integers in $O(1)$ time.

B.2. Figures

Figure 1 shows an example figure generated directly from \LaTeX using TikZ. You can also include an external file, which should be either a PDF, PNG or a JPEG file, using

```
\begin{figure}\centering
\includegraphics[width=.8\linewidth]{figures/example.jpg}
\caption{Explanation of the figure}
\label{fig:figure-label}
\end{figure}
```

It is often convenient to size figures relative to the `\linewidth`, and to use cropped files. Usually it is best to let \LaTeX do the figure placement. Do not waste time to manually force figures into your preferred location.

Table 1: This is an example table.

first row; first column	first row; second column
second row; first column	second row; second column

both columns, no lines

B.3. Tables

Tabular data is, unfortunately, often a bit clumsy in \LaTeX .

Table 1 on page x gives a very basic example.

`\reftab{tab:table}` on `\refpage{tab:table}` gives a very basic example.

Many designers frown on the overuse of lines in tables, but this is personal taste.

B.4. Algorithms

There are several packages to display algorithms. A popular and up to date package is `algorithm2e`, which shines with a very detailed and good documentation (the syntax is not that easy to learn, though).

When including algorithms, use **pseudocode** that gives a high-level description of the algorithm, instead of source code. Source code itself is better attached as a CD to the thesis.

Algorithm 1 is an example algorithm using this package.

Algorithm 1: Deep Thought

Input: The question of life, the universe, and everything

Output: The answer

Data: Data banks, the size of a small city

```

1 for  $x \in [1 \dots 7.500.000]$  do
2   | Sleep 1 Year;                               /* Pretend to think */
3 end
4  $x \leftarrow 6 \cdot 7 + (e^{i\pi} + 1) \int_x \int_y \int_z \log \sqrt[z]{x^y} dx dy dz;$  /* Compute the answer */
5 return  $x$ ;

```

B.5. Theorems and Proofs

You can define environments for theorems, propositions, etc. easily:

```
\newtheorem{theorem}{Theorem}
\newtheorem{proposition}{Proposition}
\newtheorem{corollary}{Corollary}
\begin{theorem}{Theorem}\label{thm:theorem}
There is at least one theorem in this appendix.
\end{theorem}
\begin{proof}
See Theorem~\ref{thm:theorem}
\end{proof}
```

Theorem 1. *Theorem There is at least one theorem in this appendix.*

Proof. See Theorem 1

□

B.6. Utility Commands

This templates includes some additional commands:

<code>\TODO{things to do}</code>	produces:	TODO: things to do
<code>\reffig{fig:label}</code>		Figure 1
<code>\reftab{tab:label}</code>		Table 1
<code>\refsec{sec:label}</code>		Section B.6
<code>\refalg{alg:label}</code>		Algorithm 1
<code>\refpage{sec:label}</code>		page xi

For example, to get: “Table 1 on page x shows an example.” we only need:
`\reftab{tab:table}` on `\refpage{tab:table}` shows an example.

C. Tips & Tricks for L^AT_EX

Nice guides and examples for L^AT_EX can be found, for example, here:

<http://latex-cookbook.net/> and <http://texample.net/>

D. References and Literature

D.1. Citing Literature

It is *very* important that you give appropriate credit to the sources that you used.

Try to avoid direct quotes, but it is preferable if you rephrase the information *in your own words*. But even then you must give credit to the source of the information appropriately.

References should use BibTeX. Pay extra attention to completeness and correctness.

Important: **Check all your references carefully!**

Different kinds of references (books, journal articles, papers in conference proceedings, ...) require different fields (page numbers, conference title, journal title, ... see, e.g., [Wik18]).

Warning: Google Scholar, Mendeley, and even publishers often generate flawed entries: missing authors, wrong authors, missing journal titles, wrong syntax, ...²

Some basic style guides:

- Reference ideas (and quotes), but not every single word.
- Chapter and page information is only needed for longer works such as books.
- It is common to postfix references: “We use the LOAD graph [SG16].”
- When used as subject, use the names: “Spitz and Gertz [SG16] show that ...”
With a single author, use singular, with multiple authors, use plural.
- The sentence should be grammatically correct without any bracketed term.
...but this violation of the rule is usually accepted: “as used in [SG16].”

Pro-Tip: *First* add the references to bibtex, *then* write the sentences and insert the reference *immediately*. Do not postpone references, because reconstructing them days later is *much* more work than typing “\cite{abc}” as you write the text. It does not harm to add a bibtex entry and never use it – only used entries are included.

²Tip: DBLP <http://dblp.org/> has good BibTeX-entries for much of the literature in our domain.

D.2. Citing Web Pages

With web pages you have to give the title, authors, URL, **access date**, and – if possible – an **archived copy**, because web pages can change at any time.

The appropriate BibTeX type for a web page is `@misc`.

If you *really* cannot find a *better* source than Wikipedia (primary sources are always preferable to secondary sources such as Wikipedia, and with Wikipedia you do not know who the real author of a statement is), then at least use the “cite this page” function.

But even this function produces flawed BibTeX entries, so *fix the errors*.

Example with a *permanent* URL (but unspecific authors): “Wikipedia contributors. *BibTeX* – *Wikipedia, The Free Encyclopedia*. [Online; accessed 12-April-2018]. 2018. URL: <https://en.wikipedia.org/w/index.php?title=BibTeX&oldid=830618636>”

D.3. Citing Images

Images are usually considered to be stand-alone, copyrighted, works.

§60a UrhG allows the use of figures and excerpts in education *to a large extent*,³ but:

§63 UrhG: “When a work or part of a work [...] is duplicated or distributed, **you must always clearly indicate the source**.” (name the author, *not just give an URL!*)

Obey copyright laws (i.e., German Urheberrecht). It’s not fun, but necessary.

³You need to look up the legislation details yourself, e.g., on <https://www.bmbf.de/de/neues-urheberrechtsgesetz-fuer-die-wissenschaft-4431.html> (German)

References

- [SG16] Andreas Spitz and Michael Gertz. “Terms over LOAD: Leveraging Named Entities for Cross-Document Extraction and Summarization of Events.” In: *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016*. 2016, pp. 503–512. doi: 10.1145/2911451.2911529.
- [Wik18] Wikipedia contributors. *BibTeX — Wikipedia, The Free Encyclopedia*. [Online; accessed 12-April-2018]. 2018. URL: <https://en.wikipedia.org/w/index.php?title=BibTeX&oldid=830618636>.