

Titel der Arbeit

Vorname Nachname

Matrikelnummer

17. Mai 2024

Proseminar der AG Data Mining, Informatik, TU Dortmund

An dieser Stelle bitte eine kurze Zusammenfassung der Ergebnisse (nicht jedoch der Struktur) des Dokuments. Was erfährt man, wenn man ihr Dokument gelesen hat, was ist der neue Beitrag? Das schreibt man sinnvoller am Ende, wenn die eigentliche Arbeit abgeschlossen ist, und man den gesamten Inhalt kennt.

1 Einleitung

Sinnvollerweise fängt die Arbeit mit einer Einleitung an, in der Sie den Kontext der Arbeit beschreiben. In folgenden Kapiteln werden in der Regel zunächst Vorarbeiten besprochen, dann die neue Methodik, Experimente und ein Schluss.

Beachten Sie den üblichen Schreibstil bei *wissenschaftlichen* Arbeiten. Zu [guter wissenschaftlicher Praxis](#) und Redlichkeit gehört insbesondere, dass Sie keine fremden Inhalte als ihre eigenen ausgeben. Wenn Sie sich auf fremde Beiträge beziehen, benennen Sie die Quellen und schreiben Sie so, dass *nicht* der Eindruck entsteht, dies wären ihre eigenen Leistungen. Insbesondere experimentelle Ergebnisse aus anderen Arbeiten sind – leider – oft mit Vorsicht zu genießen.

Das in der akademischen Literatur übliche „man“ (im Englischen einfacher „we“) ist in der Regel als „Autorinnen, Autoren, Leserinnen und Leser“ zu interpretieren, und beinhaltet nicht die Autorinnen und Autoren einer Quelle. Die erste Person ist nur üblich, wenn die Subjektivität einer Aussage betont werden soll. Verwenden Sie bei Quellen bevorzugt die echten *Namen*. Mit dem `\citeauthor{}` Kommando können Sie dabei peinliche Tippfehler vermeiden.

Allgemeine Hinweise zum Schreiben von Studien- und Abschlussarbeiten finden Sie unter anderem in folgenden beiden Lehrbüchern, auf die Sie über VPN zugreifen können sollten:

M. Deininger, H. Lichter, J. Ludewig und K. Schneider, *Studienarbeiten*, 6. Aufl. vdf Hochschulverlag AG, Mai 2017, ISBN: 978-3-72-813812-5. Adresse: <https://elibrary.utb.de/doi/book/10.5555/9783728138125>

R. Voss, *Wissenschaftliches Arbeiten: ... leicht verständlich!* utb GmbH, Juni 2022, ISBN: 978-3-83-858812-4. DOI: [10.36198/9783838588124](https://doi.org/10.36198/9783838588124)

2 Zielgruppe

Die Zielgruppe einer Arbeit sind in der Regel Studierende im gleichen Studienverlauf wie Sie. Inhalte aus dem Pflichtstudium oder vorausgesetzten Modulen können Sie daher als bekannt voraussetzen, und für diese sind auch meist keine Quellenangaben notwendig (bspw. QuickSort können Sie ohne Quellenangabe nennen). Bei einer Seminararbeit können Sie auch Grundlagen aus dem Rahmenthema als bekannt voraussetzen. Bei Abschlussarbeiten können Sie auch Vertiefungsmodule der Arbeitsgruppe als Grundwissen ansehen, dass sie nicht wiederholen müssen.

Versuchen Sie ihre Arbeit so zu erstellen, dass sie einen *Mehrwert* für diese Zielgruppe liefert, und für diese *gut verständlich* ist. Der Betreuer hingegen wird in der Regel die Originalquellen kennen oder bevorzugen. Um den Rückgriff auf diese zu vereinfachen, sollten sie wie üblich zitiert sein.

3 Sprache

Laden Sie die passenden Sprachpakete mit `babel`. Das letztgenannte Sprachpaket legt dabei die Hauptsprache der Arbeit fest. Mit der korrekten Sprache werden bestimmte Überschriften (bzw. „Literaturverzeichnis“) richtig erzeugt, der Anführungszeichenstil angepasst, und insbesondere die *korrekte Silbentrennung* aktiviert.

You can temporarily switch the language to obtain correct hyphenation in other languages. Simply use an other language environment. This also affects quotes and citations, e.g., “Deiningner *et al.* [1]”.

Es mag ihnen auf den ersten Blick einfacher *erscheinen*, auf Deutsch zu schreiben. Da aber ein Großteil der Quellen und Fachbegriffe auf Englisch ist, kommt dann schnell ein ziemlicher Kauderwelsch heraus, auch bekannt als „Denglisch“. Des Weiteren neigen die meisten auf Deutsch zu „Schachtelsätzen“, worunter die Lesbarkeit stark leidet, während es auf Englisch den meisten leichter fällt, einen einfacheren Satzbau zu verwenden. Viele der Schwierigkeiten mit geschlechtergerechter Sprache entfallen ebenfalls, wenn Sie auf Englisch schreiben, insbesondere wenn man auf den Plural (they statt he/she) ausweichen kann. Eine sogenannte „Generalklausel“ erfüllt nicht das Ziel, sprachinduzierte Stereotype aufzulösen. Für Sie besteht aber aktuell wohl keine Verpflichtung zum Gendern, lediglich eine Empfehlung, und daher wird dies nicht bei der Benotung berücksichtigt.¹ Es empfiehlt sich dennoch, dies zu üben.

Wenn Sie später einmal eine wissenschaftliche Arbeit international veröffentlichen wollen – beispielsweise Teile ihrer Masterarbeit – wird dies auf Englisch erfolgen müssen, damit sie auch gelesen wird. Daher ist es eine gute Übung, bereits bei Seminar- und Bachelorarbeiten auf Englisch zu schreiben. Zudem gibt es für Englisch meist die besseren Werkzeuge bspw. zur Grammatikprüfung.

Auf jeden Fall sollten Sie in ihrem Editor eine *Rechtschreibprüfung* installieren und einsetzen! Es empfiehlt sich auch, schwierige Sätze mit einer Grammatikprüfung zu validieren. Auch Verwandte, Bekannte und Mitstudierende können helfen, Fehler zu finden. Rechtschreib- und Grammatikfehler hinterlassen immer einen schlechten ersten Eindruck.

¹Siehe: <https://gleichstellung.tu-dortmund.de/themen/geschlechtergerechte-sprache/>

4 L^AT_EX-Vorlage

In diesem Seminar wird erwartet, dass die zur Verfügung gestellte L^AT_EX-Vorlage verwendet wird. Mit Hinblick auf die Anforderungen der Disziplin, insbesondere Formelsatz und Quellenangaben, ist L^AT_EX derzeit Mittel der Wahl. Auch wenn manche Studierende eine Konvertierung von Markdown nach L^AT_EX bevorzugen, so gewinnt man damit nur sehr wenig – die in Markdown möglichen Formatierungen sind auch in L^AT_EX kein großer Mehraufwand, und der Umweg über Markdown bietet hier nahezu keine Vorteile.

5 Literatur

Legen Sie bereits *beim Lesen* eine Literaturdatenbank an, sinnvollerweise im Bibtex-Format. Saubere Bibtex-Einträge finden Sie für einen Großteil der Informatik-Literatur auf <https://dblp.org/>. Das Open-Source Werkzeug JabRef (<https://jabref.org/>) kann bequem aus DBLP importieren. Nichts ist ärgerlicher, als wenn Sie eine verwendete Quelle selbst nicht mehr wiederfinden!

Kopieren Sie nicht aus anderen Werken. Wörtliche Zitate sind in der Informatik eher unüblich. Wenn Sie wirklich etwas aus einer Quelle übernehmen (bspw. eine Grafik), fügen Sie **immer als erstes die Quellenangabe** ein, dann vergessen Sie das nicht. Wörtliche Übernahmen setzen Sie *sofort in Anführungszeichen*. Auch ein automatisches Paraphrasieren ist *nicht empfehlenswert*. Sie sollen zeigen, dass Sie die Quelle verstanden haben, nicht dass Sie Strg+C kennen. Plagiate sind nicht akzeptabel! Geben Sie dazu die Inhalte mit ihren eigenen Worten wieder – dann kann man beim genauen Lesen bspw. nachvollziehen, warum Sie zu einem anderen Ergebnis gekommen sind.

biblatex (mit biber) ist eine modernere und konfigurierbare Alternative zu dem klassischen bibtex, das Sie in vielen Anleitungen noch finden. Deswegen setzen wir in dieser Vorlage auf biblatex.

Kontrollieren Sie das Literaturverzeichnis! Hier finden sich oft peinliche Fehler. So war in einem Literaturverzeichnis durch einen nicht korrigierten Parsing-Fehler von Google Scholar die fehlerhafte Autorenangabe „Petra *Stein-Sven* Vollnhals“ statt der korrekten Angabe „Petra Stein und Sven Vollnhals“.

Mit den Kommandos `\textcite{}` und `\parencite{}` können Sie Quellen zitieren. Dabei ist bei der Verwendung als Subjekt in einem Satz die Verwendung von Beispiel: Deininger u. a. [1] bzw. am Ende eines Satzes [2].

Das Kommando `\fullcite{}` erlaubt auch eine komplette Angabe im Fließtext (siehe oben), was sich aber nur für herausragende Quellen anbietet (bspw. die Hauptquelle einer Seminar- oder Bachelorarbeit).

Quellenangaben mit `\textcite{}` kommen an die richtige Stelle im Satzbau, oftmals ist dies das Subjekt am Anfang des Satzes. Ein Satz sollte ohne die Klammern immer noch Sinn ergeben, daher sollte `\parencite{}` nur verwendet werden, wenn es grammatikalisch nicht mitgelesen werden muss. Quellenangaben die sich nur auf einen Begriff beziehen direkt hinter diesen. Beziehen sie sich jedoch auf den ganzen Satz, so kommen sie in der Regel vor das Satzzeichen. Manche Autor*innen nutzen Quellenangaben nach dem letzten Satzzeichen, wenn sie sich auf den gesamten Absatz beziehen.

Internetquellen sind mit Vorsicht zu handhaben, denn die Webseiten können sich jederzeit ändern. Deswegen müssen Sie das Zugriffsdatum mit angeben. Idealerweise sichern sie eine Kopie der Seite, entweder bei ihren Unterlagen oder mit einem Archivdienst wie <https://web.archive.org/> oder <https://archive.is/>. Zudem lassen sich Autor*innen nicht immer klar identifizieren. Sind Autor*innen genannt, beispielsweise bei einem Zeitungsartikel, dann *müssen* Sie diesen aber auch in der Quellenangabe nennen. Wikipedia *hat* ein Zitationswerkzeug, mit dem Sie einen Link zu einer spezifischen Version der Seite erstellen können, kann aber dennoch Fehler in den BibTeX-Einträgen enthalten.

6 Aufzählungen

Da dies keine PowerPoint-Präsentation ist, werden sie selten Aufzählungen benötigen.

- einfach mehrere `\items` erstellen
- auch nicht komplizierter als Markdown...

Natürlich gibt auch nummerierte Aufzählungen.

1. Eins
2. Zwei

Aus Platzgründen ist es oftmals aber geschickt, Aufzählungen im Text eingebettet zu verwenden: (1)~Eins, (2)~Zwei.

7 Schriftsatz und Typografie

Weder im Deutschen noch im Englischen macht man Leerzeichen von Satzzeichen. Nach einem Satzzeichen wird ein größerer Abstand verwendet, dies macht aber \LaTeX automatisch. Um das zu verhindern – bspw. bei Abkürzungen – können Sie entweder `.{}` verwenden, oder ein geschütztes Leerzeichen `~`. Geschützte Leerzeichen sind des weiteren hilfreich (1) vor Quellenangaben, d.h., `~\parencite{}`, (2) bei um einen Zeilenumbruch innerhalb von logischen Einheiten zu verhindern (bspw. in 3~Dimensionen.), oder um unschöne Umbrüche am Ende eines Satzes oder in einer Aufzählung zu verhindern (bspw. (1) vor Quellenangaben). Leere Klammern `{}` sind auch bei Makros oft hilfreich um Ambiguitäten zu vermeiden. Möchten Sie den Zeilenumbruch bspw. in einem Algorithmennamen vermeiden, so können Sie mit `\mbox{QuickSort}` einen Umbruch verhindern.

Unterschieden wird zwischen einem Bindestrich (data-set), der mathematischen Operation ($a - b$), einem Gedankenstrich (–), und einem langen Strich (—). Um Text zu betonen nutzen Sie vorwiegend `\emph{}` statt **Fettdruck**, während Unterstriche als veraltet gelten. Mit `\enquote{}` können Sie korrekte Anführungszeichen setzen, „auch wenn diese ‚geschachtelt‘ sind.“ Apostrophen sind im Deutschen fast nur als Auslassungszeichen richtig, nicht bei jedem Genitiv oder gar Plural.

8 Tabellen

In Tabelle 1 sehen Sie ein Beispiel für eine Tabelle als Fließobjekt auf der Seite. Aufgrund der Platzierungslogik von \LaTeX ist es meist geschickt, die Tabelle am Anfang des Abschnitts in dem sie verwendet werden soll einzufügen.

Tabelle 1: Tabellen haben ihre Beschreibung in der Regel oben (Bilder unten), und es werden nur sparsam Linien verwendet.

Links	Rechts	ausgerichtete Zahlen
links	rechts	123.45 ms
links	rechts	3.45 s

Mit den Buchstaben `bth` können Sie die Platzierung steuern. Die „here“ (`h`) Platzierung ist dabei eher ungewöhnlich. Fließobjekte am Anfang

(t) oder Ende (b) einer Seite spart in der Regel Platz aufgrund der nötigen Abstände vor und nach eines Fließobjektes. Und gerade bei größeren Tabellen kann eine „here“ Platzierung einen Seitenbruch nötig machen. Bei einer echten Verwendung im Text kann es passender sein, nur ein tabbing einzusetzen. Versehen Sie jede Tabelle mit einem `\label{tab:uniqueid}`, dann erhalten Sie mit `\ref{tab:uniqueid}` eine automatische Referenz darauf.

9 Bilder und Grafiken

Bilder können in \LaTeX mit dem Kommando `\includegraphics` eingebunden werden. Dabei bieten sich für `pdflatex` die Dateiformate `.png`, `.jpg` und `.pdf` (für Vektorgrafiken) an. Bilder werden in der Regel als Fließobjekt in einer `\figure`-Umgebung platziert, im Gegensatz zu Tabellen steht hierbei die `\caption` aber unterhalb. Die Verwendung von `\label` und `\ref` ist analog.

Achten Sie unbedingt auf eine ausreichende Bildqualität! Ein Bildschirmfoto sieht oftmals gedruckt auf einmal sehr schlecht aus, insbesondere ein Bildschirmfoto von Text. Wenn ihr System dabei Subpixel-Antialiasing einsetzt können unerwünschte Farbschlieren entstehen. Überlegen Sie, ob sie nicht das gewünschte Material (1) als Text, (2) als Vektorgrafik, oder (3) hochauflösend erstellen können. Bspw. können Sie im Browser den Zoom erhöhen bevor Sie einen Screenshot einer Webseite machen.

Mit den Paketen `tikz` und `pgfplots` lassen sich auch komplexe Grafiken und Diagramme in \LaTeX selbst „programmieren“. Ein wesentlicher Vorteil dabei ist, dass dann die Schriften im Dokument konsistent sind, und man beispielsweise mathematische Formeln problemlos in Beschriftungen verwenden kann.

10 Code und Pseudocode

Für Programmcode und Pseudocode gibt es zahlreiche Lösungen. Probieren Sie aus, welcher Ansatz ihnen am Besten gefällt. Das Python-Programm `pygments` kann Quellcode in verschiedenen Sprachen formatieren, und auch als \LaTeX -Fragment ausgeben. Das Paket `minted` ruft automatisch `pygmentize` auf, was aber die Laufzeit erhöht und die Option `-shell-escape` benötigt. Daher kann es

elegant sein, Code-Fragmente automatisch mit einer klassischen `Makefile` und `pygmentize` zu übersetzen und lediglich bei `\input` einzubinden. Das Paket `listings` scheint hier etwas einfacher zu nutzen zu sein.

Für Pseudocode sind auch die Pakete `algorithmicx`, `algpseudocode` und `algorithm2e` populär. Wählen Sie eine Lösung die ihnen optisch und von der Syntax her gefällt.

11 Formeln

\LaTeX ist insbesondere bekannt für den hervorragenden Formelsatz.

$$\sum_{i=1}^N \frac{1}{i} = 1 + \frac{1}{2} + \frac{1}{3} + \dots$$

ist die harmonische Reihe. Für $N \rightarrow \infty$ gilt

$$\sum_{i=1}^{\infty} \frac{1}{i} \rightarrow \infty.$$

Durch Verwendung der `align*` Umgebung sind die Formeln mit `&` aneinander ausgerichtet.

Der Mathemodus sollte dabei nicht für *Kursivdruck* (*Kursivdruck?*) missbraucht werden. Der Ausdruck `abc` ist gleichbedeutend mit `$a b c$`, das Produkt der drei Variablen *a*, *b* und *c*. Text und Funktionsnamen in Formeln können je nach Kontext mit `\text{natürliche Sprache}`, `\mathrm{name}` oder `\operatorname{operator}` formatiert werden (vgl. `funktion_1`, `funktion_1`, `funktion_1`, `funktion_1` – der Unterschied zwischen den letzten drei hängt von den eingestellten Schriften sowie dem Kontext ab). Für häufige Notation bietet es sich dabei an, eigene Makros zu definieren:

```
\newcommand{\argmin}{\operatorname*{arg\,min}}
\newcommand{\argmax}{\operatorname*{arg\,max}}
\newcommand{\norm}[1]{\lVert {#1} \rVert}
\newcommand{\sprod}[2]{\left\langle {#1}, {#2} \right\rangle}
```

um bequemer folgendes Ergebnis zu erhalten:

$$\arg \min_x \|x\|$$

$$\|x\|^2 = \langle x, x \rangle$$

Makros erlauben es insbesondere, die Notation nachträglich noch leicht anzupassen.

12 Kürzen

Spätestens wenn Sie einmal eine eigene Publikation schreiben wollen, werden Sie eine Seitenzahlbegrenzung erfüllen müssen. Typischerweise haben Konferenzen eine maximale Seitenzahl (begrenzte früher die Druckkosten, heute begrenzt das hauptsächlich den oft enormen Aufwand der Gutachter*innen).

Ein paar Tricks mit denen man \LaTeX -Dokumente oft kürzer bekommt:

- Platzierung von Grafiken und Tabellen: wenn diese im Fließtext „schwimmen“, so kommt in der Regel oben *und* unten ein Abstand hinzu. Die – übliche! – Platzierung oben oder unten auf der Seite (mit [tb]) spart oft eine Zeile.
- Suchen Sie nach kurzen Zeilen am Ende von Paragraphen. Oft kann bereits das Entfernen oder Tauschen von einem Wort ausreichen, um eine Zeile zu sparen
- Ungünstige Spalten- und Seitenumbrüche. Viele Stile vermeiden Witwen- und Waisenzeilen. Das Kürzen um eine Zeile kann durch Kaskadierungseffekte dann schon einen erheblichen Unterschied machen, wenn so mehrere Zeilen und ein Paragraphenabstand auf die vorherige Seite rutschen. Achten Sie auf vergrößerte Abstände bei Überschriften und Formeln, sowie am Ende einer Spalte, diese weisen oft darauf hin, dass ein Kürzen an dieser Stelle mehrere Zeilen von der nächsten Seite holen kann.
- Formeln sind schwer umzuberechnen. Hier kann es manchmal helfen mit kleinen Änderungen etwas Platz zu sparen, bspw. mit `\nolimits` und `\textstyle`:

$$\sum_{x=1}^N \frac{1}{x} \text{ vs. } \sum_{x=1}^N \frac{1}{x} \text{ vs. } \sum_{x=1}^N \frac{1}{x}$$

Literatur

- [1] M. Deininger, H. Lichter, J. Ludewig und K. Schneider, *Studienarbeiten*, 6. Aufl. vdf Hochschulverlag AG, Mai 2017, ISBN: 978-3-72-813812-5. Adresse: <https://elibrary.utb.de/doi/book/10.5555/9783728138125>.
- [2] R. Voss, *Wissenschaftliches Arbeiten: ... leicht verständlich!* utb GmbH, Juni 2022, ISBN: 978-3-83-858812-4. DOI: [10.36198/9783838588124](https://doi.org/10.36198/9783838588124).