

BACHELORARBEIT

Performance – Optimierung von Datenbanken

vorgelegt am 26. März 2022
Daniel Freire Mendes

Erstprüferin: Prof. Dr. Stefan Sarstedt
Zweitprüfer: Prof. Dr. Olaf Zukunft

**HOCHSCHULE FÜR ANGEWANDTE
WISSENSCHAFTEN HAMBURG**

Department Informatik
Berliner Tor 7
20099 Hamburg

Zusammenfassung

Der Arbeit beginnt mit einer kurzen Beschreibung ihrer zentralen Inhalte, in der die Thematik und die wesentlichen Resultate skizziert werden. Diese Beschreibung muss sowohl in deutscher als auch in englischer Sprache vorliegen und sollte eine Länge von etwa 150 bis 250 Wörtern haben. Beide Versionen zusammen sollten nicht mehr als eine Seite umfassen. Die Zusammenfassung dient u. a. der inhaltlichen Verortung im Bibliothekskatalog.

Abstract

The thesis begins with a brief summary of its main contents, outlining the subject matter and the essential findings. This summary must be provided in German and in English and should range from 150 to 250 words in length. Both versions combined should not comprise more than one page. Among other things, the abstract is used for library classification.

Inhaltsverzeichnis

Abbildungsverzeichnis	III
Tabellenverzeichnis	IV
1 Überblick	1
1.1 Einführung in Benchmarks	1
1.2 Measures	2
1.3 Tools	3
1.3.1 Einführung	3
1.3.2 Kurze Einführung in die Tools	4
2 Allgemeines	7
3 Die einzelnen Teile der Arbeit	8
3.1 Titelseite	8
3.1.1 Titelseite	9
3.1.2 Abstract	9
3.2 Inhaltsverzeichnis und andere Verzeichnisse	9
3.3 Gliederungsebenen	9
3.4 Literaturverzeichnis	10
3.5 Anhang	10
3.6 Eigenständigkeitserklärung	11
4 T_EXnik und Typographie	12
4.1 Verwendung der Vorlage	12
4.2 Textformatierung	13
4.3 Mehrsprachiger Text	14
4.4 PDF-Ausgabe	14
4.5 Verweise	14
4.5.1 Querverweise	14
4.5.2 Zitate und Literaturverweise	14
4.6 Mathematische Formeln	15

4.7	Elemente in Gleitumgebungen	16
4.7.1	Tabellen	17
4.7.2	Abbildungen	17
4.8	Code	19
Literatur		21
Anhang		22

Abbildungsverzeichnis

4.1	Schlecht: Rastergrafik	18
4.2	Besser, aber noch nicht gut: Vektorgrafik	18
4.3	Erstes TIKZ-Beispiel	19
4.4	Zweites TIKZ-Beispiel	20

Tabellenverzeichnis

3.1	Dateien der Vorlage	8
4.1	Durchfallquoten Mathematik	17

1 Überblick

1.1 Einführung in Benchmarks

Benchmarks dienen dazu, praktisch und effektiv zu untersuchen, wie sich ein System unter Last verhält. Die wichtigsten Erkenntnisse, die man aus Benchmarks gewinnen kann, sind die Probleme und Fehler, die man systematisch dokumentieren und nach Priorität abarbeiten sollte. Das Ziel von Benchmarks ist die Reduzierung und Bewertung von unerwünschtem Verhalten sowie die Analyse, wie sich das System derzeit und unter simulierten, zukünftigen, anspruchsvolleren Bedingungen verhalten könnte.

Es gibt zwei verschiedene Techniken für Benchmarks. Die erste zielt darauf ab, die Applikation als Ganzes zu testen (full-stack). Dabei wird nicht nur die Datenbank getestet, sondern die gesamte Applikation, einschließlich des Webserver, des Netzwerks und des Applikationscodes. Der Ansatz dahinter ist, dass ein Nutzer genauso lange auf eine Abfrage warten muss, wie das gesamte System benötigt. Daher sollte diese Wartezeit so gering wie möglich sein. Es kann dabei vorkommen, dass MySQL nicht immer das Bottleneck ist.¹

Full-Stack-Benchmarks haben jedoch auch Nachteile. Sie sind schwieriger zu erstellen und insbesondere schwieriger korrekt einzurichten. Wenn man lediglich verschiedene Schemas und Abfragen in MySQL auf ihre Performance testen möchte, gibt es sogenannte Single-Component-Benchmarks. Diese analysieren ein spezifisches Problem in der Applikation und sind deutlich einfacher zu erstellen. Ein weiterer Vorteil besteht darin, dass nur ein Teil des gesamten Systems getestet wird, wodurch die Antwortzeiten kürzer sind und man schneller Ergebnisse erhält.

Wenn bei Benchmarks schlechte Designentscheidungen getroffen werden, kann dies zu einer falschen Interpretation des Systems führen, da die Ergebnisse nicht die Realität widerspiegeln. Die Größe des Datensatzes und des Workloads muss realistisch sein. Idealerweise verwendet

¹Gemeint ist ein Engpass beim Transport von Daten oder Waren, der maßgeblichen Einfluss auf die Abarbeitungsgeschwindigkeit hat. Optimierungsversuche an anderer Stelle führen oft nur zu geringen oder gar keinen messbaren Verbesserungen der Gesamtsituation. (<https://martinvogel.de/lexikon/bottleneck.html>)

man einen Snapshot² des tatsächlichen produktiven Datensatzes. Gibt es keine Produktionsdaten, sollten die Daten und der Workload simuliert werden, da realistische Benchmarks komplex und zeitaufwendig sein können.

Häufige Fehler beim Durchführen von Benchmarks sind unter anderem, dass nur ein kleiner Teil der tatsächlichen Datensatzgröße verwendet wird und die Datensätze unkorrekt gleichmäßig verteilt sind. In der Realität können Hotspots auftreten. Bei zufällig generierten Werten kommt es hingegen häufig zu unrealistisch gleichmäßig verteilten Datensätzen. Ein weiterer Fehler besteht darin, dass man beim Testen einer Anwendung nicht das tatsächliche Benutzerverhalten nachstellt. Wenn gleiche Abfragen in einer Schleife ausgeführt werden, muss man außerdem auf das Caching achten, da sonst falsche Annahmen über die Performance getroffen werden können. Zudem wird oft die Warmmachphase des Systems vollständig ignoriert. Kurze Benchmarks können schnell zu falschen Annahmen über die Performance des Systems führen.

Um verlässliche Ergebnisse zu erhalten, sollte ein Benchmark ausreichend lange laufen, um den stabilen Zustand des Systems zu beobachten, insbesondere bei Servern mit großen Datenmengen und viel Speicher. Dabei ist es wichtig, so viele Informationen wie möglich zu erfassen und sicherzustellen, dass der Benchmark wiederholbar ist, da unzureichende oder fehlerhafte Tests wertlos sind. Außerdem ist es wichtig, die Ergebnisse in einem Diagramm darzustellen, da auftretende Phänomene sonst anhand einer tabellarischen Darstellung nicht erkannt werden können.

1.2 Measures

- **Durchsatz (Throughput):** Der Durchsatz ist die Anzahl an Transaktionen pro Zeiteinheit. Er ist standardisiert, und Datenbankanbieter versuchen, diesen zu optimieren. Meistens werden Transaktionen pro Sekunde (oder manchmal pro Minute) als Einheit verwendet.
- **Antwortzeiten (Latenz):** Die Antwortzeit misst die gesamte Zeit, die für eine Abfrage benötigt wird. Diese kann, abhängig von der Applikation, in Mikrosekunden (μ s), Millisekunden (ms), Sekunden oder Minuten angegeben werden. Von dieser Zeit können aggregierte Antwortzeiten wie Durchschnitt, Maximum, Minimum und Perzentile abgeleitet werden. Das Maximum ist oft eine weniger sinnvolle Metrik, da es sich nicht gut wiederholen lässt. Daher nutzt man eher Perzentile bei den Antwortzeiten. Wenn beispielsweise das 95. Perzentil der Antwortzeit bei 5 ms liegt, bedeutet dies, dass mit einer Wahrscheinlichkeit von 95 % die Abfrage in weniger als 5 ms abgeschlossen ist.

²Snapshots bestehen größtenteils aus Metadaten, die den Zustand Ihrer Daten definieren, und sind keine vollständige Duplikation der Daten auf Ihrer Festplatte. Snapshots werden häufig für Test-/Entwicklungsaufgaben verwendet. (<https://www.rubrik.com/de/insights/what-is-a-snapshot-backup>)

- **Nebenläufigkeit (Concurrency):** Die Nebenläufigkeit auf dem Webserver lässt sich nicht zwangsläufig auf den Datenbankserver übertragen. Eine genauere Messung der Gleichzeitigkeit auf dem Webserver besteht darin, zu bestimmen, wie viele gleichzeitige Anfragen zu einem bestimmten Zeitpunkt ausgeführt werden. Es kann auch geprüft werden, ob der Durchsatz sinkt oder die Antwortzeiten steigen, wenn die Gleichzeitigkeit zunimmt. Beispielsweise benötigt eine Website mit „50.000 Benutzern gleichzeitig“ vielleicht nur 10 oder 15 gleichzeitig laufende Abfragen.
- **Skalierbarkeit (Scalability):** Skalierbarkeit ist wichtig für Systeme, die ihre Performance unter unterschiedlich starken Workloads beibehalten müssen. Ein ideales System würde doppelt so viele Abfragen beantworten (Throughput), wenn doppelt so viele „Arbeiter“ versuchen, die Aufgaben zu erfüllen. Die meisten Systeme sind jedoch nicht linear skalierbar und zeigen Leistungseinbußen, wenn die Parameter variieren.

1.3 Tools

1.3.1 Einführung

- [Webhoster Wissen - MySQL Benchmark mit Sysbench](#)
- [Sysbench GitHub Repository](#)

Als Haupttool, um Benchmarktests durchzuführen, habe ich mich für Sysbench entschieden. Sysbench ist ein Open-Source-Tool, das ein skriptfähiges, multi-threaded Benchmark-Tool, das auf LuaJIT basiert. Es wird auch hauptsächlich für Datenbankbenchmarks verwendet, kann jedoch auch dazu eingesetzt werden, beliebig komplexe Arbeitslasten zu erstellen, die keinen Datenbankserver erfordern. Dabei werden Tests auf verschiedenen Systemressourcen, wie CPU, Speicher, I/O und Datenbanken wie MySQL verwendet.

Im Zuge der Recherchearbeit habe ich mir auch andere Benchmarking-Tools betrachtet, wie z.B. Benchbase ([GitHub](#)) oder mybench ([GitHub](#)). Die größten Vorteile von Sysbench habe ich in der Skriptfähigkeit und Flexibilität gesehen. D.h. dass ich benutzerdefinierte Benchmarks schneller und unkompliziert erstellen kann. Außerdem hat sich Sysbench als de facto Standard im Bereich der Datenbankbenchmarks etabliert ([mybench](#)). Dadurch stehen eine breite Nutzerbasis und viele verfügbare Ressourcen zur Verfügung. Im Vergleich zu den anderen Tools bietet allerdings Sysbench eine weniger präzise Steuerung der Ergebnisrate und der Transaktionen. Außerdem haben Tools wie mybench die Möglichkeit, in Echtzeit umfassende Visualisierungen darzustellen. Damit können Metriken live in einem Diagramm angezeigt werden ([mybench Live Monitoring](#)). Dieses Feature ist sicherlich hilfreich, aber in meinem Fall habe ich abgewogen und bin zu dem Entschluss gekommen, dass die einfachere

Bedienung für mich der ausschlaggebende Grund, neben dem Fakt, dass Sysbench der de facto Standard ist.

Da die Graphen aber trotzdem eine entscheidende Rolle bei der Analyse darstellen, werde ich das Tool Gnuplot ([GitHub](#)) dafür benutzen, um die Werte nach der Durchführung des Benchmarks zu visualisieren.

1.3.2 Kurze Einführung in die Tools

Zunächst müssen die beiden Tools installiert werden. Auf einem Mac erfolgt dies über:

```
1 brew install sysbench
2 brew install gnuplot
```

Mit der Hilfe von Homebrew. Des Weiteren muss der MySQL-Server korrekt gestartet sein und eine Datenbank erstellt werden mit:

```
1 CREATE DATABASE sbtest;
```

Um einfach Testdaten in die Datenbank einzufügen, kann dieser Befehl verwendet werden:

```
1 sysbench --db-driver=mysql --mysql-host=localhost --mysql-user=YOUR_USER --mysql-
  password=YOUR_PASSWORD --mysql-db=sbtest oltp_insert --tables=10 --table-size
  =100000 prepare
```

YOUR_USER und YOUR_PASSWORD müssen entsprechend ersetzt werden.

Codeblock 1.1: Sysbench Script

```
1 #!/bin/bash
2
3 # File Paths
4 OUTPUT_FILE="output/sysbench_output.csv"
5 RAW_RESULTS_FILE="output/sysbench.log"
6 GNUPLOT_SCRIPT="plot_sysbench.gp"
7
8 # Connection parameters
9 DB_HOST="localhost"
10 DB_USER="root"
11 DB_PASS="password"
12 DB_NAME="sbtest"
13 TABLES=10
```

```

14 TABLE_SIZE=10000
15 DURATION=10
16
17 echo "Preparing the database..."
18 sysbench oltp_read_write --db-driver=mysql --mysql-host=$DB_HOST --mysql-user=
    $DB_USER --mysql-password=$DB_PASS --mysql-db=$DB_NAME --tables=$TABLES --table-
    size=$TABLE_SIZE prepare > $RAW_RESULTS_FILE 2>&1
19 echo "Database prepared."
20
21 echo "Running benchmark..."
22 sysbench oltp_read_write --db-driver=mysql --mysql-host=$DB_HOST --mysql-user=
    $DB_USER --mysql-password=$DB_PASS --mysql-db=$DB_NAME --tables=$TABLES --table-
    size=$TABLE_SIZE --time=$DURATION --threads=1 --report-interval=1 run >>
    $RAW_RESULTS_FILE 2>&1
23
24 echo "Time (s),Threads,TPS,QPS,Reads,Writes,Other,Latency (ms,95%),Err/s,Reconn/s" >
    "$OUTPUT_FILE"
25
26 # Extract the relevant lines and format as CSV
27 grep '^[' ' $RAW_RESULTS_FILE | while read -r line; do
28     time=$(echo $line | awk '{print $2}' | sed 's/s//')
29
30     threads=$(echo $line | awk -F 'thds: ' '{print $2}' | awk '{print $1}')
31     tps=$(echo $line | awk -F 'tps: ' '{print $2}' | awk '{print $1}')
32     qps=$(echo $line | awk -F 'qps: ' '{print $2}' | awk '{print $1}')
33
34     read_write_other=$(echo $line | sed -E 's/.*\((r\|w\|o: ([0-9.]+)\)/([0-9.]+)
    \)/([0-9.]+)\).*\/\1,\2,\3/')
35     reads=$(echo $read_write_other | cut -d',' -f1)
36     writes=$(echo $read_write_other | cut -d',' -f2)
37     other=$(echo $read_write_other | cut -d',' -f3)
38
39     latency=$(echo $line | awk -F 'lat \\\(ms,95%\\\): ' '{print $2}' | awk '{print $1
    }')
40     err_per_sec=$(echo $line | awk -F 'err/s: ' '{print $2}' | awk '{print $1}')
41     reconn_per_sec=$(echo $line | awk -F 'reconn/s: ' '{print $2}' | awk '{print $1
    }')
42
43     echo "$time,$threads,$tps,$qps,$reads,$writes,$other,$latency,$err_per_sec,
    $reconn_per_sec" >> "$OUTPUT_FILE"
44 done
45

```

```

46 echo "Benchmark complete. Results saved to $OUTPUT_FILE."
47
48 # Generate the Gnuplot graph
49 echo "Generating plot..."
50 gnuplot $GNUMPLOT_SCRIPT
51 echo "Plot generated."
52
53 echo "Cleaning up..."
54 sysbench oltp_read_write --db-driver=mysql --mysql-host=$DB_HOST --mysql-user=
    $DB_USER --mysql-password=$DB_PASS --mysql-db=$DB_NAME --tables=$TABLES cleanup
    >> $RAW_RESULTS_FILE 2>&1
55 echo "Database cleanup complete."

```

Codeblock 1.2: Gnuplot Script

```

1 set datafile separator ","
2 set title "Benchmark Results: TPS, Latency, Queries, and More"
3 set xlabel "Time (s)"
4 set ylabel "Values"
5 set grid
6 set key outside
7 set terminal pngcairo enhanced font 'Arial,10'
8 set output "/Users/danielmendes/Desktop/Bachelorarbeit/Ausarbeitung/Tools/Output/
    sysbench_output_plot.png"
9 set yrange [0:∞]
10
11 # Plot each attribute on its own line
12 plot "/Users/danielmendes/Desktop/Bachelorarbeit/Ausarbeitung/Tools/Output/
    sysbench_output.csv" using 1:2 title "Threads" lt 1 lc rgb "black" with lines, \
13     "" using 1:3 title "TPS" lt 2 lc rgb "green" with lines, \
14     "" using 1:4 title "QPS" lt 3 lc rgb "blue" with lines, \
15     "" using 1:5 title "Reads" lt 4 lc rgb "red" with lines, \
16     "" using 1:6 title "Writes" lt 5 lc rgb "orange" with lines, \
17     "" using 1:7 title "Other" lt 6 lc rgb "purple" with lines, \
18     "" using 1:8 title "Latency (ms)" lt 7 lc rgb "cyan" with lines, \
19     "" using 1:9 title "Err/s" lt 8 lc rgb "magenta" with lines, \
20     "" using 1:10 title "Reconn/s" lt 9 lc rgb "brown" with lines

```

2 Allgemeines

Diese Vorlage zur Verwendung mit \LaTeX ¹ kann für die Formatierung Ihrer Bachelorarbeit verwendet werden, ist jedoch keine verbindliche Vorgabe; bitte stimmen Sie sich hier mit Ihrer betreuenden Erstprüferin bzw. Ihrem betreuenden Erstprüfer ab.

Den aktuellen Stand dieser Vorlage entnehmen Sie bitte dem Datum auf der Titelseite.

Das Textsatzsystem \LaTeX ist für die Erstellung druckfertiger wissenschaftlicher Arbeiten unabhängig von deren Umfang hervorragend geeignet, weil es explizit dafür konzipiert ist. Im Allgemeinen sollten Sie damit bessere Ergebnisse als mit Textverarbeitungsprogrammen wie WORD oder PAGES erzielen. Allerdings handelt es sich nicht um ein mit der Maus zu bedienendes [WYSIWYG](#)-Programm und erfordert eine gewisse Einarbeitungszeit. Die Verwendung von \LaTeX für eine Abschlussarbeit ist daher nicht unbedingt zu empfehlen, wenn Sie das System erst kurz vor dem Schreiben der Arbeit zum ersten Mal benutzen.

In diesem Sinne ist das vorliegende Dokument auch explizit *nicht* als Einführung in \LaTeX gedacht, sondern setzt voraus, dass Sie schon ausreichende Kenntnisse mitbringen. Die können Sie beispielsweise durch die Lektüre von Büchern wie (Voß, [2021](#)) oder (Schlosser, [2021](#)) erwerben.

¹In diesem Dokument wird den üblichen Gepflogenheiten entsprechend nicht zwischen dem zugrundeliegenden Textsatzsystem \TeX und dem weitverbreiteten Makropaket \LaTeX für dieses System unterschieden, obwohl das rein technisch falsch ist.

3 Die einzelnen Teile der Arbeit

In diesem Kapitel geht es um die wesentlichen Teile, die eine Abschlussarbeit haben sollte. Die Dateien, auf die hier Bezug genommen wird, werden in in Tabelle 3.1 vorgestellt und in Kapitel 4 näher erläutert. Die Vorlage wurde absichtlich in relativ viele Dateien zerlegt, um zu demonstrieren, wie man mit dem Befehl `\includeonly` nur Teile der Arbeit kompilieren kann, um Zeit zu sparen. Vor dem Druck muss natürlich das komplette, aus allen Dateien bestehende Dokument kompiliert werden (ggf. mehrfach), damit alle Querverweise (siehe Abschnitt 4.5.1) korrekt sind.

Tabelle 3.1: Dateien der Vorlage

Dateiname	Zweck
VorlageBA.tex	Hauptdatei
defs.tex	Laden von Paketen und Setzen von Optionen
title.tex	Titelseite und Abstract
toc.tex	Inhaltsverzeichnis und optionale Verzeichnisse
chap1.tex bis chap3.tex	erstes, zweites und drittes Kapitel
appendix.tex	Literaturverzeichnis, Anhang und Eigenständigkeitserklärung
demo.bib	exemplarische Bibliographie
HAW_Marke_RGB_300dpi.jpg	HAW-Logo für Titelseite
bitmap.png, vector.pdf und euler.py	Beispieldateien

3.1 Titelei

Mit dem Begriff *Titelei* bezeichnet man im Buchwesen den Teil eines Buches, der dem eigentlichen Inhalt vorangestellt ist. In dieser Vorlage dient der Absatz, den Sie gerade lesen, im Wesentlichen aber nur als Vorwand, eine weitere Unterebene einzufügen.

3.1.1 Titelseite

Die Titelseite befindet sich in der Datei `title.tex` und ist der [Vorlage](#) der HAW nachempfunden. Sie müssen natürlich in dieser Datei den Titel, die Namen und das Datum anpassen. Außerdem müssen die [Hausschriften](#) der HAW [installiert](#) sein.

3.1.2 Abstract

Nach der Titelseite folgt der Abstract, der sich ebenfalls in der Datei `title.tex` befindet. Hier ersetzen Sie den Beispieltext durch eine möglichst aussagekräftige Zusammenfassung Ihrer Arbeit.

3.2 Inhaltsverzeichnis und andere Verzeichnisse

Nach dem Abstract (siehe Abschnitt [3.1.2](#)) folgt das Inhaltsverzeichnis, das von \LaTeX automatisch erzeugt wird. Exemplarisch wird in dieser Vorlage auch gezeigt, wie man Abbildungs- und Tabellenverzeichnisse generieren könnte; siehe dazu die Datei `toc.tex`. Das ist in der Regel aber nur dann nötig, wenn es in Ihrer Arbeit sehr viele Abbildungen bzw. Tabellen gibt. Bei Bedarf können auch Verzeichnisse von Codeblöcken, mathematischen Formeln oder anderen Objekten erzeugt werden; mehr dazu in (Voß, [2021](#), Kap. 12). Hier gilt aber ebenfalls, dass das normalerweise nicht nötig sein sollte. Auch ein Glossar oder ein Abkürzungsverzeichnis ist in einer Bachelorarbeit eher unüblich. Sprechen Sie ggf. mit Ihrer Erstprüferin oder Ihrem Erstprüfer ab, was wirklich gebraucht wird.

3.3 Gliederungsebenen

Eine Bachelorarbeit sollte in der Regel mit maximal drei Gliederungsebenen auskommen. In der Dokumentenklasse der Vorlage entspricht das den Befehlen `\chapter`, `\section` und `\subsection`, die auch für eine automatische Aufnahme der jeweiligen Abschnitte ins Inhaltsverzeichnis sorgen. Weitere Gliederungsebenen verringern typischerweise die Lesbarkeit des Dokuments. Falls Sie das bei Ihrer Arbeit trotzdem für nötig halten, sprechen Sie es vorher mit der Erstprüferin bzw. dem Erstprüfer ab.

Auf jeder Gliederungsebene sollte es mindestens zwei Abschnitte derselben Hierarchie geben, da ansonsten eine Gliederung auf dieser Ebene sinnlos wäre. Wenn Sie also einen Abschnitt 2.3.1 haben, dann muss es mindestens einen weiteren Abschnitt 2.3.2 geben; anderenfalls sollte alles unter 2.3 stehen.

Außerdem sollten einzelne Abschnitte einen Umfang haben, der die entsprechende Gliederung rechtfertigt. Dieser Text, in dem Abschnitte meistens nur aus wenigen Sätzen bestehen, ist dafür ein schlechtes Beispiel!¹ Es handelt sich allerdings auch um eine Vorlage und nicht um eine Bachelorarbeit ...

In der Vorlage ist jedes Kapitel in eine eigene Datei ausgelagert. Beachten Sie, dass der Befehl `\include` grundsätzlich eine neue Seite anfängt. Unterabschnitte von Kapiteln müssen daher mit `\input` eingefügt werden, wenn sie in separaten Dateien liegen sollen.

3.4 Literaturverzeichnis

Die Vorlage enthält ein exemplarisches Literaturverzeichnis, das nach dem sogenannten APA-Standard formatiert ist und das als Beispiele einige Bücher, einen Fachartikel und eine Internetquelle umfasst. Das Literaturverzeichnis befindet sich am Ende der Arbeit vor dem Anhang. Falls Ihre Erstprüferin oder Ihr Erstprüfer ein anderes Format wünscht, ist das mit dem verwendeten Paket [BibLaTeX](#) ohne große Probleme zu realisieren, siehe z. B. (Voß, 2021, Kap. 13). Wichtig ist, dass das Literaturverzeichnis vollständig ist und Sie nur die Quellen angeben, die Sie im Rahmen Ihrer Bachelorarbeit wörtlich zitiert oder sinngemäß wiedergegeben haben. Literatur, die Sie lediglich zur Vorbereitung genutzt haben, gehört nicht in das Literaturverzeichnis.²

Wenn Sie die von Ihnen verwendeten Texte nach dem Muster der Datei `demo.bib` eingeben, wird das Literaturverzeichnis automatisch einheitlich dargestellt und alphabetisch sortiert. Die Literaturverwaltung [CITAVI](#), für die es eine Hochschullizenz gibt, kann Daten im sogenannten BibTeX-Format ausgeben. (Das ist das in `demo.bib` verwendete Format.)

Verwenden Sie wissenschaftliche Quellen. (Hinweis: Wikipedia gilt nicht als wissenschaftliche Quelle.) Wenn die Angabe von Internetquellen unumgänglich ist, muss das Datum des letzten Aufrufs angegeben werden. Ein Beispiel dafür finden Sie in der Vorlage.

Wie man zitiert, wird in Abschnitt [4.5.2](#) gezeigt.

3.5 Anhang

Falls zu Ihrer Arbeit Datenreihen, Quelltexte, transkribierte Interviews oder weitere ergänzende Informationen gehören, dann gehören diese in einen Anhang ganz am Ende. Ob ein Anhang

¹Darum wirken die Abstände und Seitenaufteilungen in dieser Vorlage auch vergleichsweise unruhig. Bei längeren Texten wird es besser aussehen.

²Wenn Sie die Voreinstellungen der Vorlage verwenden, werden aber ohnehin nur die Quellen ins Literaturverzeichnis übernommen, die auch zitiert werden.

notwendig ist und welchen Umfang er haben sollte, sollten Sie vorab mit Ihrer Erstprüferin oder Ihrem Erstprüfer absprechen. Häufig ist es sinnvoller, die Daten auf einem Datenträger der Arbeit beizulegen.

Die Vorlage enthält einen exemplarischen Anhang in der Datei `appendix.tex`.

3.6 Eigenständigkeitserklärung

Die letzte Seite der Arbeit ist die Eigenständigkeitserklärung, die Sie in der Datei `appendix.tex` finden. Tragen Sie hier den Titel Ihrer Arbeit sowie den Ort und das Datum ein. Alle gedruckten Exemplare werden dann oberhalb des Datums von Ihnen unterschrieben.

4 T_EXnik und Typographie

Die grundsätzliche Philosophie von L^AT_EX ist, dass sich die Autoren auf den Inhalt konzentrieren und um das Erscheinungsbild des Textes keine großen Gedanken machen sollen. Die typographischen „Entscheidungen“, die das System bzw. die Dokumentenklasse trifft, sind in der Regel sinnvoll und führen zu besseren Ergebnissen als manuelle Eingriffe ungeübter Benutzer.

Wenn Sie sich dabei erwischen, dass Sie Abstände, Positionen, Größen oder andere Parameter manipulieren, dann machen Sie sehr wahrscheinlich etwas falsch. (Eine Ausnahme sind Dinge wie die Titelseite, die einer Vorgabe nachempfunden werden sollen.)

In diesem Kapitel werden einige *Best Practices* aufgeführt, die besonders die Leser aufmerksam studieren sollten, die noch nicht so erfahren im Umgang mit L^AT_EX sind. Auf diverse typische Fehler im Umgang mit L^AT_EX weist das Video <https://youtu.be/OSzs9K6-fRQ> hin, aber es ist kein Ersatz für ein einführendes Buch.

4.1 Verwendung der Vorlage

Die Vorlage setzt ein installiertes T_EX-System wie M^IK_T_EX oder T_EX LIVE voraus. Sie ist für das Kompilieren mit L^UA_T_EX gedacht (und wird daher mit P_DF_T_EX nicht ohne entsprechende Änderungen funktionieren). Die Bibliographie wird mit B_IB_ER bearbeitet.

Es ist nicht möglich, in diesem Rahmen für jede Entwicklungsumgebung die richtigen Optionen anzugeben. Daher nur Anmerkungen zu drei weitverbreiteten Tools:¹

- Für T_EXstudio muss man in der Konfiguration als *Standardcompiler* die Option LuaLaTeX auswählen und als *Standard Bibliographieprogramm* die Option Biber.
- Für T_EXworks benötigt man einen Durchlauf mit der Einstellung LuaLaTeX und dann einen mit der Einstellung Biber. Danach lässt man das Dokument noch einmal mit der Einstellung LuaLaTeX kompilieren und kann in dieser Einstellung bleiben. Einen weiteren Biber-Durchlauf zwischendurch benötigt man nur dann, wenn in der Bibliographie etwas geändert wurde.

¹Die ExpertInnen wissen natürlich, wie man so etwas automatisieren kann.

- Wenn Sie [LATEXMK](#) benutzen, dann können Sie die Vorlage einfach mit

```
latexmk -lualatex -bibtex VorlageBA.tex
```

kompilieren und alles sollte ohne weiteren Eingriff funktionieren.

Nach diesen Schritten sollte das erzeugte PDF exakt so aussehen wie das, das Sie gerade lesen. Die Vorlage besteht aus mehreren Dateien, die in Tabelle 3.1 auf Seite 8 aufgeführt sind. Für einen kurzen Text wie diesen könnte man zur Not auch mit einer Datei auskommen, aber für umfangreichere Dokumente ist eine Aufteilung nach diesem Muster empfehlenswert. Für L^AT_EX müssen die Quelldateien (also die mit der Endung `.tex`) UTF-8-kodiert sein. Die Vorlage ist als Gerüst für Ihre Arbeit gedacht und als Vorschlag zu betrachten. In den meisten Fällen gibt es alternative Methoden, um zu vergleichbaren Ergebnissen zu kommen. Wenn Sie sich gut mit L^AT_EX auskennen, spricht nichts dagegen, beispielsweise andere Pakete, andere Befehle oder andere Optionen zu verwenden.

Die Dateien sind mit ausführlichen Kommentaren versehen. Viele Informationen zur Verwendung der Vorlage finden Sie nur in diesen Kommentaren und *nicht* in dem Text, den Sie gerade lesen!

Die Datei, die kompiliert werden muss und die die anderen lädt, ist `VorlageBA.tex`. Diese Datei sollten Sie zunächst umbenennen (z. B. in `BANameVorname.tex`), damit das erzeugte PDF-Dokument keinen generischen Namen hat, sondern Ihnen zugeordnet werden kann. Wenn Sie die anderen Dateien umbenennen, dann müssen in den entsprechenden Aufrufen von Befehlen wie `\include` oder `\addbibresource` auch die Namen angepasst werden.

4.2 Textformatierung

In der Vorlage sind, abgesehen von der Titelseite, die freien Schriftarten [Libertinus](#) für Fließtext und [Anonymous Pro](#) für Code in einer Größe von 12 Punkt eingestellt. Diese Schriften sind im Druck und auf dem Bildschirm gut lesbar und aufeinander abgestimmt. Außerdem bietet diese Kombination volle Unterstützung für mathematischen Formelsatz. Wenn Sie kein Experte für Typographie sind, sollten sie keine Zeit damit verschwenden, andere Schriftsätze auszuprobieren. Der Zeilendurchschuss ist etwas großzügiger bemessen, als man es für den Druck eines Buches machen würde. Auch diesen Parameter sollten Sie nicht ohne triftigen Grund ändern.

4.3 Mehrsprachiger Text

Die Vorlage ist für Text in deutscher Sprache angelegt, in dem auch englische Zitate vorkommen können. Dafür wird das Standardpaket `BABEL` eingesetzt. Damit die automatische Silbentrennung sowie einige andere Details richtig funktionieren, muss \LaTeX „wissen“, welche Sprache gerade verwendet wird. Beispiele dafür, wie man dem System den Wechsel der Sprache mitteilt, findet man in den Dateien `chap3.tex` und `title.tex`. Weitere Sprachen kann man in der Datei `defs.tex` hinzufügen.

This paragraph only serves to demonstrate that "quotation marks" will be treated correctly depending on the choice of language.

4.4 PDF-Ausgabe

Wenn Sie die Vorlage wie vorgesehen verwenden, werden durch das Paket `HYPERRREF` automatisch Lesezeichen zur bequemen Navigation in den gängigen PDF-Programmen gesetzt. Außerdem können (und sollten) dem Dokument auch Metadaten zugewiesen werden. Dies geschieht am Ende der Datei `defs.tex`. Dort müssen Sie Ihren Namen und den Titel Ihrer Arbeit eingeben. Ebenfalls am Ende von `defs.tex` wird erklärt, wie Sie `HYPERRREF` so umkonfigurieren, dass für die gedruckte Version Ihrer Arbeit Links nicht farbig dargestellt werden.

4.5 Verweise

4.5.1 Querverweise

Wenn Sie für Querverweise auf Abschnitte, Formeln, Seiten, Abbildungen usw. die Standardbefehle `\label`, `\ref`, `\eqref` und `\pageref` benutzen, sorgt \LaTeX automatisch dafür, dass alles richtig nummeriert wird und im PDF zusätzlich Links gesetzt werden. In der Vorlage finden Sie diverse Beispiele für den Einsatz dieser Befehle. Sie sollten Querverweise auf keinen Fall manuell in der Form „siehe Kapitel 3“ oder „auf Seite 42“ eintippen!

4.5.2 Zitate und Literaturverweise

Bei Verweisen auf die Literatur werden immer Autor und Jahr angegeben, typischerweise in der Form (Weitz, 2021b). Wörtliche Zitate werden zusätzlich durch Anführungszeichen oder durch kursive Schrift gekennzeichnet und mit der Seitenzahl versehen. Es folgen ein paar Beispiele:

- (i) Man kann die Leibnizformel mit zahlentheoretischen Methoden herleiten. Das ist allerdings relativ umständlich (Weitz, 2021a).
- (ii) Wie Weitz in (2021a) schreibt, kann man die Leibnizformel mit zahlentheoretischen Methoden herleiten. Das sei allerdings relativ umständlich
- (iii) Das Fazit des Autors nach Abwägen aller Alternativen: "The crux of the biscuit is the apostrophe" (Zappa, 1974, S. 42).²
- (iv) Zappa kommt in (1974) schließlich und endlich zum Ergebnis: „The crux of the biscuit is the apostrophe“ (S. 42).
- (v) Man kann die Leibnizformel mit zahlentheoretischen Methoden herleiten. Das ist allerdings relativ umständlich.³

Die Versionen (i) und (iii) sind am übersichtlichsten und werden empfohlen. Die Alternative (v) – Quellenangaben in den Fußnoten – ist für Bachelorarbeiten in unserem Department eher unüblich und sollte nur nach Absprache verwendet werden.

4.6 Mathematische Formeln

Mathematischer Formelsatz ist die Spezialität von \LaTeX . Während man allerdings Texte in \LaTeX fast so wie in einem Textverarbeitungsprogrammen schreiben kann, muss man für Formeln eine eigene Sprache lernen. Einen „Crashkurs“ für diese mathematische Syntax, die man inzwischen auch in WORD und vielen anderen Programmen verwenden kann, bietet das Video <https://youtu.be/7ovgNXRiJ6g>. (Siehe auch das am Anfang von Kapitel 4 erwähnte Video zu typischen Fehlern.) Auch hier gilt aber die Anmerkung, dass ein Lehrbuch die bessere Quelle sein wird.

Die Vorlage ist so eingerichtet, dass eingerückte Formeln nicht zentriert werden, sondern wie beispielsweise

$$\int_a^b g(t) dt = \int_a^b \Re(g(t)) dt + i \int_a^b \Im(g(t)) dt \quad (4.1)$$

etwas versetzt am linken Rand anfangen. Diese Einstellung kann man beim Aufruf des Befehls `\documentclass` ändern, indem man die Option `fleqn` entfernt.

²Hier werden die Anführungszeichen des fremdsprachlichen Zitates verwendet, während im nächsten Beispiel deutsche Anführungszeichen um den fremdsprachlichen Text herum benutzt werden. Insbesondere für kurze Zitate in anderen Sprachen gibt es dafür keine einheitlichen Regeln. Sie sollten aber zumindest konsistent vorgehen und im gesamten Dokument nur eine von beiden Alternativen verwenden.

³Weitz, 2021a.

Im mathematischen Kontext ist es üblich, dass Formeln wie Satzbestandteile der normalen Sprache behandelt werden. Das hat insbesondere Auswirkungen auf die Interpunktion. Zum Beispiel besagt die verallgemeinerte Kontinuumshypothese, dass für jede Ordinalzahl α die Gleichung $\aleph_\alpha = \beth_\alpha$ gilt. In der Sprache der Zermelo-Fraenkel-Mengenlehre heißt das

$$\forall \alpha \in \text{ON} \quad 2^{\aleph_\alpha} = \aleph_{\alpha+1}.$$

Dieses Beispiel sollte nur demonstrieren, dass hier an das Ende der Formel ein Punkt gehört, weil der Satz dort endet. Abgesetzte Formeln wie die gerade gezeigte sollten normalerweise nicht nummeriert werden. Eine Ausnahme davon sind natürlich Formeln wie (4.1), auf die später mit `\eqref` verwiesen wird.

Die Vorlage lädt automatisch die Erweiterung [AMS-L^AT_EX](#), die viele hilfreiche Erweiterungen für mathematische Ausdrücke enthält. Ohne dieses Paket wäre es beispielsweise wesentlich aufwendiger, Kettenbrüche wie

$$e = \sum_{k=0}^{\infty} \frac{1}{k!} = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = 2 + \frac{1}{1 + \frac{1}{2 + \frac{2}{3 + \frac{3}{4 + \frac{4}{5 + \ddots}}}}}$$

darzustellen.

4.7 Elemente in Gleitumgebungen

Tabellen und Abbildungen sollten grundsätzlich mit einer Beschriftung (`\caption`) versehen sein. Dazu gehört auch eine Quellenangabe, wenn es sich um Daten handelt, die Sie übernommen haben. Außerdem sollten solche Objekte immer in eine sogenannte Gleitumgebung wie `table` oder `figure` gesetzt werden, damit $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ für eine richtige Platzierung sorgen kann und die Objekte ggf. in Verzeichnissen (siehe Abschnitt 3.2) erfasst werden können. Viel mehr zu diesem Thema findet man u. a. in (Voß, 2021, Kap. 11).

In der Vorlage stehen Beschriftungen für Tabellen und Codeblöcke (siehe Abschnitt 4.8) jeweils links eingerückt über den Tabellen und Beschriftungen für Abbildungen zentriert unter diesen. Achten Sie darauf, den `\caption`-Befehl entsprechend am Anfang bzw. am Ende der Gleitumgebung aufzurufen.

4.7.1 Tabellen

Für typographisch anspruchsvoll gesetzte Tabellen verwendet die Vorlage das Paket `BOOKTABS` zusammen mit der Standardumgebung `tabular`. Beispiele für Tabellen finden Sie in den Dateien `chap2.tex` und `chap3.tex`. Grundsätzlich gilt, dass Linien in Tabellen im Allgemeinen *nicht* der Übersichtlichkeit dienen, sondern bezüglich der Lesbarkeit kontraproduktiv sind. In den meisten Ratgebern zur Typographie wird insbesondere empfohlen, auf vertikale Linien komplett zu verzichten! In Tabelle 4.1 gibt es eher zu viele Linien. Die wurden dort lediglich eingefügt, um ein paar Gestaltungsmöglichkeiten von `BOOKTABS` zu demonstrieren.

Tabelle 4.1: Durchfallquoten Mathematik

<i>Fantasiewerte!</i>	Jahr			
	2018	2019	2020	2021
Sommersemester	16,5 %	15,4 %	14,3 %	13,2 %
Wintersemester	26,5 %	25,4 %	24,3 %	23,2 %
Gesamt	21,5 %	20,4 %	19,3 %	18,2 %

Rein technisch ist es möglich, mit `\includegraphics` Tabellen einzufügen, die in anderen Programmen erzeugt wurden. Das sollte aber eigentlich immer vermieden werden, weil es sehr hässlich aussieht. Siehe dazu auch die Anmerkungen in Abschnitt 4.7.2. Für Tabellen, die länger als eine Seite sind, werden spezielle Pakete wie beispielsweise `LONGTABLE` benötigt.

Es gibt diverse Tools, die Daten aus Programmen wie EXCEL einlesen und diese in \LaTeX -Syntax als Tabellen ausgeben können. Manche Programme, z. B. MATHEMATICA, können auch direkt nach \LaTeX exportieren.

4.7.2 Abbildungen

Abbildungen können mit dem Befehl `\includegraphics` eingefügt werden. Wenn es sich nicht um Fotos handelt, ist jedoch darauf zu achten, dass auf jeden Fall skalierbare Vektorgrafiken verwendet werden. Rastergrafiken (wie sie beispielsweise durch Screenshots oder Scans erzeugt werden) führen, insbesondere im Druck, fast immer zu suboptimalen Ergebnissen. Das wird etwas übertrieben in Abbildung 4.1 dargestellt.

Abbildung 4.2 zeigt im Vergleich eine Vektorgrafik. Das ist schon besser, hat aber in den meisten Fällen den Nachteil, dass die Schriften nicht mit denen Ihrer Arbeit übereinstimmen werden, was unter typographischen Aspekten sehr unschön ist.⁴ (Der zugehörige Code in der Datei `chap3.tex` zeigt nebenbei, wie man eingefügte Grafiken beschneiden kann.)

⁴Das gilt „erst recht“ für Tabellen, die aus externen Quellen eingefügt werden.

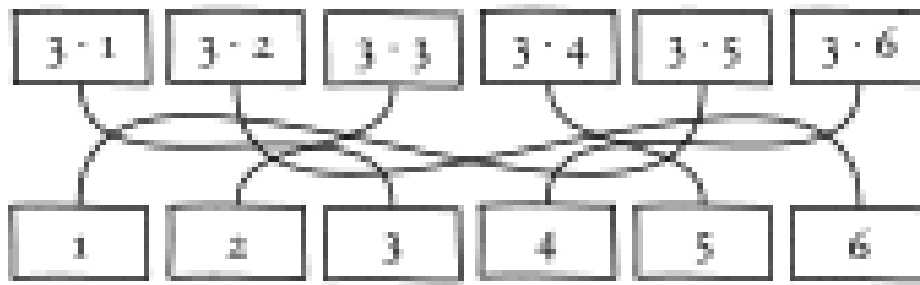


Abbildung 4.1: Ganz schlechtes Beispiel! (Quelle: Weitz, 2021a, S. 184)

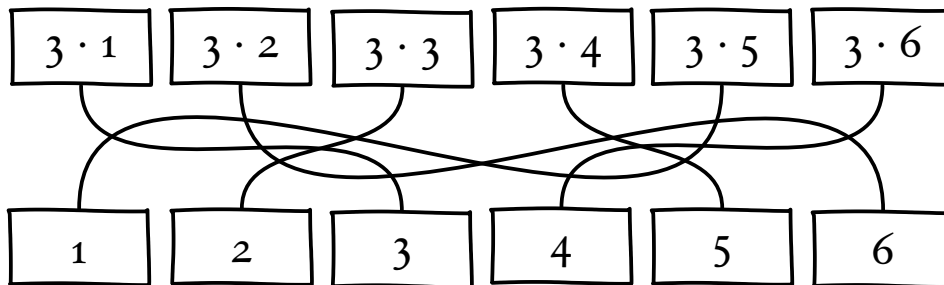


Abbildung 4.2: Besser, aber nicht optimal (Quelle: Weitz, 2021a, S. 184)

Die beste Lösung ist das Erstellen der Grafiken direkt in \LaTeX . Dazu eignet sich hervorragend das Paket `PGF/TikZ`, das nahezu unbegrenzte Möglichkeiten bietet, jedoch eine steile Lernkurve hat. Dafür ist die mitgelieferte Dokumentation allerdings auch hervorragend. Als Einführung kann man auch die Videos in der Playlist

<https://www.youtube.com/playlist?list=PLb0zKSynM2PBbpe9x6LgOZkSCL2yJAOQY>

verwenden.

Diese Vorgehensweise hat zudem den Vorteil, dass man potentiellen Problemen mit dem Urheberrecht aus dem Weg geht, weil dieses nicht auf eine nach einer Vorlage selbst erstellte Grafik anwendbar ist. Nichtsdestotrotz muss aber auch in diesem Fall die Herkunft benannt werden!

Die Abbildungen 4.3 und 4.4 zeigen Beispiele für solche Grafiken. Den zugehörigen Quellcode findet man in der Datei `chap3.tex`.

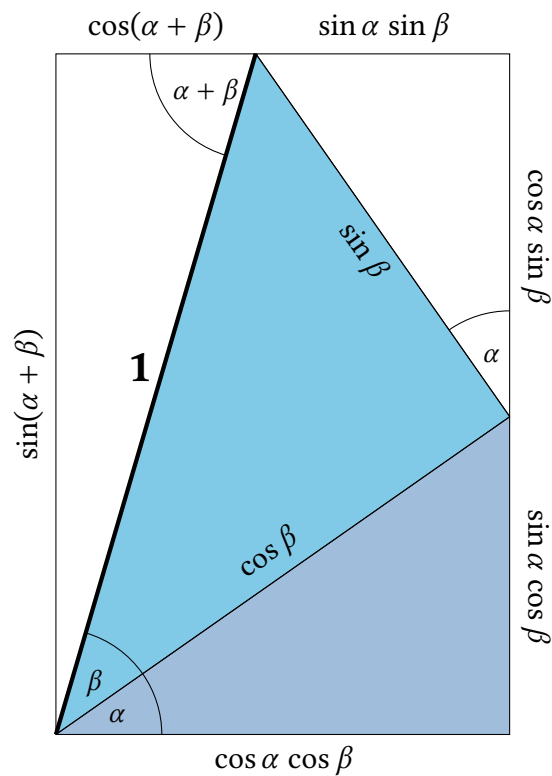


Abbildung 4.3: Mit TikZ [eigene Grafik nach (Weitz, 2021a, S. 271)]

4.8 Code

Für die Darstellung von Programmcode wie in Codeblock 4.1 wird in der Vorlage das Paket `LISTINGS` verwendet.

Codeblock 4.1: Eulersches Polygonzugverfahren

```
1 def euler(f,x0,y0,h,n):
2     x, y, result = x0, y0, [(x0,y0)]
3     for i in range(n):
4         y += f(x,y) * h
5         x += h
6         result.append((x,y))
7     return result
```

In der Datei `chap3.tex` kann man sehen, wie der Code aus einer externen Datei eingelesen wird. Durch diese Vorgehensweise kann man dafür sorgen, dass auch tatsächlich die aktuelle Version des eigenen Codes verwendet wird, und man vermeidet potentielle Fehler beim Abtippen. Man kann den Code aber auch wie in Codeblock 4.2 direkt eintippen.

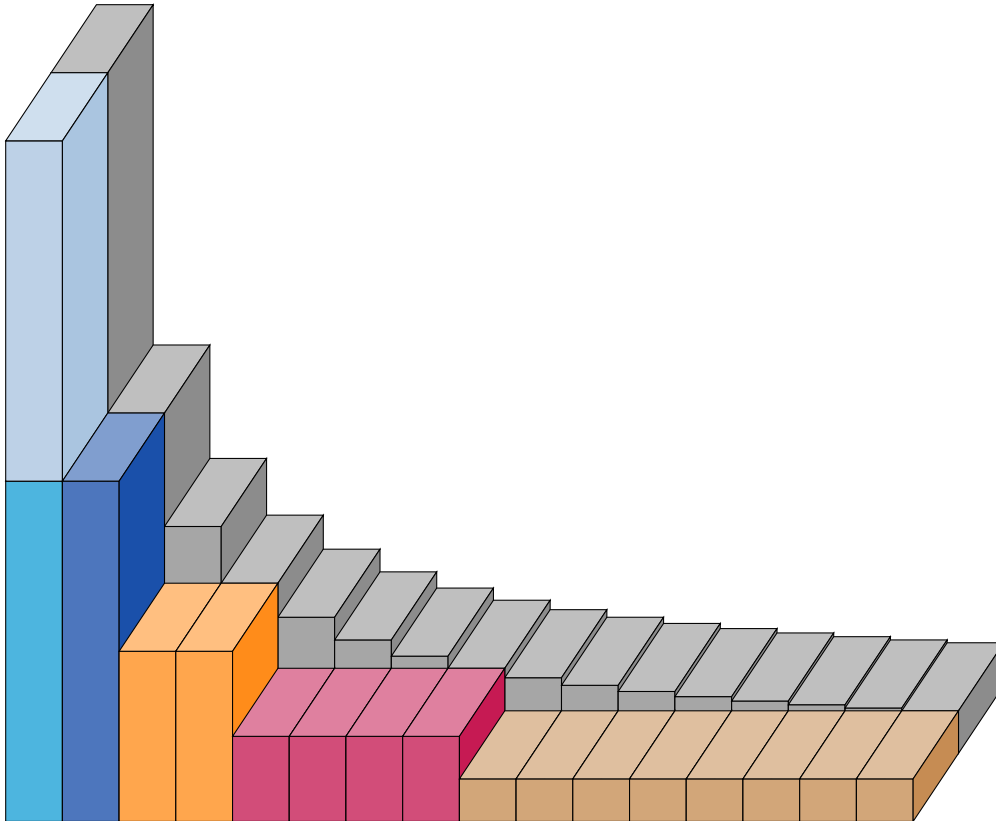


Abbildung 4.4: Auch mit TikZ [eigene Grafik nach (Weitz, 2021a, S. 533)]

Codeblock 4.2: Variationen

```

1 def kperm(L, k):
2     if k == 0:
3         return [[]]
4     return [[L[i]]+P for i in range(len(L))
5             for P in kperm(L[:i] + L[i+1:], k-1)]

```

In der Datei `defs.tex` wird exemplarisch gezeigt, wie man das Aussehen der Codeblöcke individuell gestalten kann. Es ist auch möglich, Codeblöcke wie Abbildungen und Tabellen „gleiten“ zu lassen.

Ein Verzeichnis der Codeblöcke kann mit dem Paket `LISTINGS` bei Bedarf erzeugt werden, wenn Ihre Arbeit sehr viele Codeblöcke enthält. Umfangreiche Codeblöcke gehören aber nicht in die Arbeit und auch nicht in den Anhang, sondern sollten – ggf. nach Absprache mit der Erstprüferin bzw. dem Erstprüfer – auf einem Datenträger zusammen mit der Arbeit eingereicht werden.

Literatur

- Schlosser, J. (2021). *Wissenschaftliche Arbeiten schreiben mit LaTeX: Leitfaden für Einsteiger* (7. Aufl.). mitp.
- Voß, H. (2021). *Die wissenschaftliche Arbeit mit LaTeX: unter Verwendung von LuaTeX, KOMA-Script und Biber/BibLaTeX* (2. Aufl.). Lehmanns Media.
- Weitz, E. (2021a). *Pi und die Primzahlen: Eine Entdeckungsreise in die Mathematik*. Springer.
- Weitz, E. (2021b). Hausdorff's forgotten proof that almost all numbers are normal. *Math. Semesterberichte*, 68(2), 273–282. <https://doi.org/10.1007/s00591-021-00303-w>
- Zappa, F. (1974). *Apostrophe (')*: Stink-Foot. Verfügbar 11. März 2022 unter [http://www.donlope.net/fz/lyrics/Apostrophe_\(\).html](http://www.donlope.net/fz/lyrics/Apostrophe_().html)

Anhang

Hier beginnt der Anhang. Siehe die Anmerkungen zur Sinnhaftigkeit eines Anhangs in Abschnitt [3.5](#) auf Seite [10](#).

Der Anhang kann wie das eigentliche Dokument in Kapitel und Abschnitte unterteilt werden. Der Befehl `\appendix` sorgt im Wesentlichen nur für eine andere Nummerierung.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit mit dem Titel

Viele zufällige Zahlen

selbstständig und nur mit den angegebenen Hilfsmitteln verfasst habe. Alle Passagen, die ich wörtlich aus der Literatur oder aus anderen Quellen wie z. B. Internetseiten übernommen habe, habe ich deutlich als Zitat mit Angabe der Quelle kenntlich gemacht.

Hamburg, 21. Dezember 1940