

Universität Leipzig
Medizinische Fakultät
Institut für Medizinische Informatik, Statistik und Epidemiologie

QUESTION ANSWERING AUF SNIK MIT EINEM LANGUAGE MODEL

MASTERARBEIT

zur Erlangung des akademischen Grades

Master of Science
(M. Sc.)

vorgelegt von

Paul Keller
Studiengang Medizininformatik M. Sc.

Leipzig, den 31.09.2023

AUTOR:

Paul Keller

Geboren am 23.05.1998 in Leipzig, Deutschland

TITEL:

Question Answering auf SNIK mit einem Language Model

INSTITUT:

Institut für Medizinische Informatik, Statistik und Epidemiologie
Medizinische Fakultät Universität Leipzig

REFERENT:

Prof. Dr. Alfred Winter

BETREUER:

Konrad Höffner

ABSTRAKT

DANKSAGUNG

INHALTSVERZEICHNIS

Abstrakt	iii
1 Einleitung	1
1.1 Gegenstand	1
1.2 Problemstellung	1
1.3 Motivation	3
1.4 Zielsetzung	5
1.5 Aufgabenstellung	5
1.6 Aufbau der Arbeit	5
2 Grundlagen	7
3 Stand der Forschung	9
4 Lösungsansatz	11
5 Ausführung der Lösung	13
5.1 Unterkapitel n	13
5.1.1 Unterunterkapitel n	13
6 Ergebnisse	15
7 Diskussion	17
Zusammenfassung	19
 Literatur	 21
 Appendix	
A Literaturverzeichnis-Tipps	25
A.1 BibLaTeX	25
A.2 Zitieren	25
A.2.1 Häufige Fehler	25
A.3 Zitierstile	26
B L ^A T _E X- und Schreibtipps	27
B.1 Links	27
B.2 Tabellen	27
B.3 Striche	28
B.4 Abstände und Zeilenumbrüche	29
B.5 Anführungszeichen	29
C Git-Tipps	31
C.1 GitHub und GitLab	31
D Advanced Print Tips (English)	33
D.1 One-sided vs Two-sided	33
D.2 LaTeX Warnings	33
D.2.1 Overful Hboxes	34
D.3 Color	34
D.3.1 Colored Text Elements	34
D.3.2 Colorcheck script	35

D.3.3	CMYK Color Space	35
D.4	Printing	36

ABBILDUNGSVERZEICHNIS

TABELLENVERZEICHNIS

Tabelle B.1	Diese Tabelle kann sich frei bewegen aber bevorzugt hier oder oben auf der Seite. Sie benutzt keine vertikale Linien aber verschiedene horizontale Linien mithilfe des booktabs-Paketes. 27
Tabelle B.2	Verschiedene Striche und ihr Einsatzzweck. Diese Tabelle ist so breit, dass sie rotiert dargestellt wird. Im PDF-Viewer sollte die Seite gekippt sein, sodass man den Kopf nicht drehen muss. 28

AKRONYME

EINLEITUNG

Eine Abschlussarbeit ist mit einem Projekt vergleichbar und in der Einleitung wird die Aufgabenstellung in ähnlicher Weise beschrieben wie in dem Strukturplan eines Projektes. Daher sollte die Einleitung mit Hilfe der „5-Stufen-Methode zur systematischen Strukturplanung“ (Ammenwerth u. a., 2014)¹ vorgenommen werden.

1.1 GEGENSTAND

Dieses Kapitel hilft nicht nur dem Leser, sondern auch dem Autor zu verstehen, in welchen fachlichen Kontext sich die Arbeit einordnet. Hier muss also insgesamt deutlich werden, wozu die in der Arbeit beschriebenen Arbeitsergebnisse letztlich verwendet werden sollen und welchen Nutzen sie möglicherweise bringen können.

- Welche Situation liegt vor und was soll getan werden?
- Worum geht es eigentlich?
- In welcher Welt/Domäne oder welchem Arbeitsbereich/-gebiet bewegen wir uns im Rahmen der Arbeit

1.2 PROBLEMSTELLUNG

- Warum ist die geschilderte vorliegende Situation problematisch?
- Worin bestehen die Probleme?
- Für wen ist sie problematisch?
- Welche der geschilderten Probleme sollen im Rahmen dieser Abschlussarbeit gelöst werden?

Bitte die Probleme, die in dieser Arbeit gelöst werden sollen, nummerieren und mit wenigen Sätzen zusammenfassend beschreiben:

- Problem P1: Problemname 1 mit kurzer Problembeschreibung
- Problem P2: Problemname 2 mit kurzer Problembeschreibung

Was aber sind „Probleme“ und wie kann man sie beschreiben? Die Probleme stehen im engen Zusammenhang mit den in Abschnitt 1.4 beschriebenen Zielen.

¹ Das Buch ist für Studierende der Medizininformatik an der Universität Leipzig im Moodle verfügbar. Schauen Sie in den von ihnen belegten Modulen nach.

VERMEIDEN SIE FORMULIERUNGEN WIE „ES IST NICHT BEKANNT, OB...“ ODER „ES EXISTIERT KEIN...“. Solche Formulierungen kehren in der Regel einfach das bereits angedachte Lösungsmodell um und postulieren das Fehlen der angedachten Lösung einfach als Problem. Das ist ähnlich, als wenn es in der Werbung hieße

Wenn Sie das Problem haben, dass Ihnen Aspirin fehlt,
dann kaufen Sie doch Aspirin.

Das Problem wäre dann schon gelöst, wenn Sie Aspirin gekauft haben.
Sinnvoller ist diese Aussage:

Wenn Sie das Problem haben, dass Ihnen der Kopf weh
tut, dann kaufen Sie doch Aspirin.

Es ist also bei der Problembeschreibung erforderlich, sich in die Lage dessen zu versetzen, den man mit der angedachten Lösung „beglücken“ möchte. Sein Problem ist zu ermitteln und so zu formulieren, dass er/sie das Problem wiedererkennt und dadurch geneigt ist, sich für die Lösung des Problems zu interessieren. Bei der zweiten Problembeschreibung wäre das Problem im Übrigen erst gelöst, wenn die Kopfschmerzen weg sind.

Hierzu noch ein Beispiel aus einer wissenschaftlichen Arbeit:

Problem: Es ist keine ganzheitliche Vorgehensweise bekannt, die die Entwicklung und Verbesserung von Software zur Steigerung der Motivation durch Unterhaltung in der Therapie auf strukturierte Art und Weise unterstützt.

Als Ziel wird dort formuliert:

Erstellung einer ganzheitlichen Vorgehensweise zur Entwicklung von Software zur Steigerung der Motivation durch Unterhaltung in der Therapie.

Es ist völlig unklar, wer das Problem hat, dass ihm oder ihr keine Vorgehensweise bekannt ist. Außerdem wäre das Problem schon gelöst, wenn irgendwie eine ganzheitliche Vorgehensweise bekannt würde. Tatsächlich schließt diese Arbeit auch einfach mit der Präsentation einer einheitlichen Vorgehensweise und das geschilderte Problem ist gelöst ohne noch weiter zu untersuchen, ob die Vorgehensweise irgendeinen Nutzen erbringt. Vermutlich liegt das Problem aber etwas tiefer. Es scheint doch wohl so zu sein, dass bislang völlig unbrauchbare Software entwickelt wurde und man die Hoffnung hat, durch eine ganzheitliche Vorgehensweise bessere Ergebnisse erzielen zu können. Das Problem sollte also eher so formuliert werden:

Problem: Wie aktuelle Publikationen zeigen [12-17], erfüllen die zur Zeit am Markt angebotenen Softwareprodukte zur Steigerung der Motivation durch Unterhaltung in

der Therapie nicht ihren Zweck. So wird zum Beispiel der Unterhaltungswert von 73% der Anwender als gering eingestuft und eine Motivationssteigerung für die Therapie konnte in keiner Studie nachgewiesen werden.

Dazu würde folgendes Ziel passen:

Ziel ist eine ganzheitlichen Vorgehensweise zur Entwicklung von Software zur Steigerung der Motivation durch Unterhaltung in der Therapie, durch deren Anwendung der Unterhaltungswert für die Anwender und eine Motivationssteigerung für die Therapie erreicht werden kann.

Das hier geschilderte Problem ist jetzt ein Problem von Patient:innen. Bei diesem Problem und diesem Ziel muss die Arbeit daher damit schließen, dass zumindest an einem Beispiel gezeigt wird, dass die neue Vorgehensweise tatsächlich zu gesteigertem Unterhaltungswert und zur Motivationssteigerung bei Patient:innen führt. Man könnte aber auch ‚die Latte etwas niedriger hängen‘ und sich zunächst mit folgendem Problem befassen:

Problem: Bei der Entwicklung von Softwareprodukten zur Steigerung der Motivation durch Unterhaltung in der Therapie sind übersehen viele Entwickler:innen geeignete Möglichkeiten und Spielelemente, die zur Steigerung der Motivation von Patienten eingesetzt werden können oder setzen Spielelemente für die falschen Zwecke oder an der falschen Stelle ein.

Dazu würde dann folgendes Ziel passen:

Ziel ist eine ganzheitlichen Vorgehensweise zur Entwicklung von Software zur Steigerung der Motivation durch Unterhaltung in der Therapie, die die Softwareentwickler:in ausgehend von den zu erreichenden Zielen systematisch bei der Auswahl geeigneter Spielelemente zur Steigerung von Motivation unterstützt.

In diesem Fall wird ein Problem vom Softwareentwickler:innen beschrieben und das angestrebte Ziel wird auch diese Entwickler:innen unterstützen. Hier müsste die Arbeit damit schließen, dass man zeigt, dass es bei dem Entwickeln der Software leichter wird, Elemente so einzusetzen, dass sie dem intendierten Zweck dienen.

1.3 MOTIVATION

- Warum lohnt es sich, die genannten Probleme zu lösen?
- Wer wird welchen Nutzen von dieser Abschlussarbeit haben?

- Warum ist die Arbeit wichtig?
- Wer wartet sehnlichst auf die Fertigstellung der Arbeit

Ohne ausreichende Gegenstands-, Problem- und Motivationsbeschreibung kann eine Leser:in nicht verstehen, warum z. B. eine entwickelte Software sinnvoll ist bzw. zur Lösung welches Problems sie verwendet werden soll. Gerade für die Medizinische Informatik als eine problemorientierte Disziplin ergibt sich der Wert einer Lösung, z. B. einer Software, aber vor allem daraus, ob bzw. wie weit sie ein Problem löst.

Für die Autor:in bedeutet daher eine unzureichende Gegenstands-, Problem- und Motivationsbeschreibung die Gefahr, dass sie oder er sich die zu lösende Problematik nicht ausreichend klar gemacht hat. Bei der Erstellung der Arbeit besteht dann die Gefahr, dass man möglicherweise methodisch aufregende Lösungen entwirft und realisiert, für die aber ein Problem gar nicht besteht oder die für die Lösung der tatsächlichen Probleme nicht geeignet sind. Trotz einer möglicherweise brillanten Lösung wäre dann doch eine schlechte Bewertung der Lösung und damit der Arbeit zu erwarten. Außerdem sollte sich – auch bei einer Abschlussarbeit – die Arbeit auch lohnen, d.h. es sollte genügend Motivation geben, viel Zeit und Energie zu investieren. Aus diesem Grund sollten die Kapitel 1.1 bis 1.3 ausführlich sein. Ein Umfang von weniger als drei Seiten wird in der Regel nicht ausreichen. Mit einem Augenzwinkern hier noch Motivationen, die wir nicht gerne in einer Abschlussarbeit sehen:

Warum lohnt es sich, die genannten Probleme zu lösen?	„Weil ich mein Studium endlich hinter mir haben will“, „Damit ich eine gute Note habe.“, „Weil Prof. Winter/mein Betreuer es so will.“
Wer wird welchen Nutzen von dieser Abschlussarbeit haben?	„Ich, weil ich dann endlich mit studieren fertig bin / in den Master darf.“
Warum ist die Arbeit wichtig?	„Weil mein Betreuer es möchte.“, „Weil es X noch nicht gibt.“, „Weil es X nur mit Technik Y gibt, aber nicht mit Z“, „Weil die Vorarbeiten so schlecht sind“
Wer wartet sehnlichst auf die Fertigstellung der Arbeit?	„Ich / Mein Betreuer / Prof. Winter / meine Oma“ (außer es ist eine Eigenentwicklung für speziell diese Personen)

1.4 ZIELSETZUNG

- Welche Ergebnisse werden mit dieser Abschlussarbeit angestrebt und welche der o. g. Probleme sollen damit jeweils gelöst werden?

Bitte jedes Ziel kurz oder ggf. mit Stichworten beschreiben:

- Ziel(e)/angestrebte(s) Ergebnis(se) zur Lösung von Problem P1:
 - Ziel Z1.1:
 - Ziel Z1.2:

1.5 AUFGABENSTELLUNG

- Wie sollen die o. g. Ziele erreicht werden?
- Was soll zur Erreichung der Ziele bzw. zur Schaffung der Ergebnisse getan werden?
- Welche Fragen müssen zur Erreichung der Ziele bzw. zur Schaffung der Ergebnisse beantwortet werden?

Bitte geben Sie zu jedem der o. g. Ziele mindestens zwei Aufgaben bzw. Fragen an, die bearbeitet bzw. beantwortet werden sollen. Bitte jede Aufgabe bzw. Frage kurz oder ggf. mit Stichworten beschreiben:

- Aufgaben zu Ziel Z1.1:
 - Aufgabe A1.1.1:
 - Aufgabe A1.1.2:

1.6 AUFBAU DER ARBEIT

Das Grundlagenkapitel soll den Stand der Forschung erläutern und mit Literatur belegen. Auf diesem Kapitel bauen die Erkenntnisse der Arbeit auf. Gerade in den Grundlagen wird man häufig Quellen benennen, aus denen die Aussagen letztlich stammen. Im Kapitel „Literaturverzeichnis“ dieser Vorlage wird beschrieben, wie eine Quellenangabe zu erfolgen hat.

Da sich die Medizinische Informatik mit der Lösung medizinischer Probleme befasst, sollen hier auch die Hintergründe des medizinischen Problems so dargestellt und erläutert werden, dass sie auch für Leser der Arbeit, die nicht Mediziner sind, verständlich sind.

In diesem Kapitel werden auch die Methoden erläutert, die zur Lösung der Probleme eingesetzt wurden. Stellen Sie sicher, dass hier alle, aber auch nur die Grundlagen und Methoden erläutert werden, die in der Arbeit verwendet wurden. Stellen Sie im weiteren Text der Arbeit auch sicher, dass der Leser erkennen kann, wie Sie unter Verwendung der Methoden zu Ihren Ergebnissen gekommen sind. So sollten z.B. Modellierungsmethoden nur verwendet werden, wenn die Modelle nachvollziehbar dazu genutzt werden, die Ergebnisse zu erzielen.

Bedenken Sie, dass Sie diese Arbeit zum Abschluss eines umfangreichen Studiums schreiben, das vor allem dazu diente, sie mit einem reichen Methodenrepertoire auszustatten. Wählen Sie aus den Methoden, die Sie gelernt haben, aus, benennen Sie die Methoden korrekt und wenden Sie sie an! Aber gehen sie auch kritisch mit dem um, was Ihnen gelehrt wurde. Wenn Sie feststellen, dass gelehrt Methoden ungeeignet sind, diskutieren Sie dies und suchen passendere Methoden! Wenn Sie Methoden benötigen, die nicht gelehrt wurden, suchen Sie nach passenden Methoden oder – wenn Sie nicht fündig werden – entwickeln Sie die für Ihr Problem passende Methode selbst!

Definition 1 (Beispieldefinition) *Dies ist eine Beispieldefinition.*

STAND DER FORSCHUNG

Kann auch im Grundlagenkapitel mit integriert werden! Wenn es zwei Kapitel sind, ist das Grundlagenkapitel für bekanntes „Lehrbuchwissen“ gedacht, der „Stand der Forschung“ für Vorarbeiten in Publikationen.

Grundlage für eine Abschlussarbeit ist eine gründliche Recherche im Themenumfeld. Dabei ist es ausdrücklich nicht hinreichend, mit bekannten Suchmaschinen im Internet zu recherchieren. Vielmehr wird von den Studierenden erwartet, dass sie auch referierte Veröffentlichungen (wissenschaftliche Zeitschriften (auch elektronisch), Bücher) in die Erarbeitung einbeziehen (und mit entsprechenden Quellenangaben belegen). Informationen zum Thema der Literaturrecherche finden sich in den [Hinweisen zur Literaturrecherche](#). Wir empfehlen Ihnen, Kurse zur Literaturrecherche zu belegen, z.B. beim [Academic Lab](#) oder der [Universitätsbibliothek](#). Im [Onlineportal zum Wissenschaftlichen Schreiben der Uni Leipzig](#) finden Sie wertvolle Hinweise und Übungen zu Textstruktur, Stilistik, Schreibprozess sowie Quellen und Zitaten.

BEISPIELZITIERUNGEN Höffner u. a. (2017) beschreiben ein Verfahren zur X von Y auf Basis von Z. Alternativ: X von Y lässt sich auf Basis von Z ermitteln (Höffner u. a., 2017).

LÖSUNGSANSATZ

Der Lösungsansatz ist eine kurze Beschreibung der Arbeitshypothese sowie des Vorgehens zur Lösung der in der Einleitung beschriebenen Probleme.

AUSFÜHRUNG DER LÖSUNG

Dieses Kapitel kann auch in mehrere Kapitel aufgeteilt werden, wenn das sinnvoll ist!

5.1 UNTERKAPITEL N

5.1.1 *Unterunterkapitel n*

ERGEBNISSE

Auch die Ergebnisse können in mehrere Kapitel aufgeteilt oder sogar mit der Präsentation der Ausführung verschmolzen werden. Wichtig ist, dass man zwischen Ausführung und Ergebnis unterscheiden kann.

DISKUSSION

In der Diskussion werden die Ergebnisse der Arbeit kritisch bewertet. So wäre hier z.B. darzulegen, dass man zwar eine schöne Modellierungsmethode zur Beschreibung der Kommunikation im Krankenhausinformationssystem gefunden hat, diese aber so aufwändig ist, dass kaum jemand sie benutzen wird, wenn nicht folgende weiteren Arbeiten noch durchgeführt werden.... So wird die Diskussion dann auch einen Ausblick auf die Dinge enthalten, die eigentlich noch zu erledigen sind. Gerade in einer Seminararbeit könnte hier auch die Kritik des Autors an dem stehen, was er in der Literatur zu dem zu bearbeitenden Thema hier und da gelesen hat

Dies ist eine Notiz

ZUSAMMENFASSUNG

LITERATUR

- Ammenwerth, Elske, Reinhold Haux, Petra Knaup-Gregori und Alfred Winter (2014). *IT-Projektmanagement im Gesundheitswesen*. 2. Aufl. Stuttgart, Germany: Schattauer. ISBN: 9783794530717.
- Höffner, Konrad, Franziska Jahn, Christian Kücherer, Barbara Paech, Schneider Brigel, Martin Schöbel, Sebastian Stäubert und Alfred Winter (2017). „Technical Environment for Developing the SNIK Ontology of Information Management in Hospitals“. In: *Studies in Health Technology and Informatics* 243, S. 122–126.

APPENDIX

LITERATURVERZEICHNIS-TIPPS

A.1 BIBLATEX

BibLaTeX¹ löst das alte BibTeX ab, greift aber genauso auf .bib-Dateien zu, in denen die Literaturquellen definiert sind. In dieser Vorlage ist das alte bibtex-Backend eingestellt, in `classicthesis-config.tex` kann aber auch das neuere „biber“-Backend aktiviert werden, wodurch z.B. unicode-Zeichen möglich werden und Umlaute nicht mehr escaped werden müssen. Eine .bib-Datei besteht aus einer Liste von Einträgen, die z.B. auf [Google Scholar](#) gefunden und per copy-paste in die `thesis.bib` eingefügt werden können. Diese müssen aber manuell geprüft und bei fehlenden oder inkorrekten Angaben wie Autoren oder dem Verlag korrigiert werden.

A.2 ZITIEREN

Diese Vorlage ist mit `natbib` kompatibel, sodass das neben `cite` auch die Kommandos `citet` (Zitieren im Text) und `citep` (Zitieren in Klammern) ablöst. Details stehen in der [natbib-Paketdokumentation](#).

BEISPIELZITIERUNGEN

- Höffner u. a. (2017) untersuchen verschiedene Verfahren für X.
- X ist „ein gutes Verfahren, um Y auf Basis von Z zu generieren“ (Höffner u. a., 2017).²
- In dieser Situation bietet sich auch X an, da es ein gutes Verfahren ist, um mit Z Y zu erstellen (Höffner u. a., 2017, vgl.).³

A.2.1 Häufige Fehler

- Zitierung in Fußnote
- Verwechseln von wörtlichem Zitat (Anführungszeichen), sinngemäßem Zitat („vgl.“) und anderer Aussage (weder Anführungszeichen noch „vgl.“)
- "Falsche Anführungszeichen"(immer `enquote` benutzen, dann kann dieser Fehler nicht passieren)

¹ <https://ctan.org/pkg/biblatex>

² Wörtliches Zitat

³ Sinngemäßes Zitat

A.3 ZITIERSTILE

Bei einer Publikation ist der Zitierstil normalerweise vorgegeben, aber bei unseren Abschlussarbeiten gibt es keine Vorgabe. Ein numerischer Zitierstil ist sehr kompakt und bietet sich daher an, wenn ein Seitenlimit eingehalten werden muss, aber für eine Abschlussarbeit empfehlen wir einen Autor-Jahr-Zitierstil (in dieser Vorlage bereits eingestellt), da dieser mehr Informationen vermittelt.

Manche Institute verbieten einleitenden Text außerhalb von subsections wie direkt unter diesem Kapitel hier, aber bei uns ist das ausdrücklich erwünscht.

B.1 LINKS

- [L^AT_EX @ TU Graz](#)
- [Uni Leipzig-Schreibportal](#)

B.2 TABELLEN

Titel	Zahl
Dies ist eine hässliche Beispieltabelle ohne Titel, Label, Umbruch und Float-Eigenschaft	123
Text	55.222
Noch mehr Text	111111

Titel	Zahl
Dies ist eine schöne Beispieltabelle mit Titel, Label, Umbruch, Float-Eigenschaft und korrekt eingerückten Zahlen	123
Text	55.222
Noch mehr Text	111 111

Tabelle B.1: Diese Tabelle kann sich frei bewegen aber bevorzugt hier oder oben auf der Seite. Sie benutzt keine vertikale Linien aber verschiedene horizontale Linien mithilfe des booktabs-Paketes.

B.3 STRICHE

Deutscher Name	Englischer Name	Einsatzzweck	Beispiel	Negativbeispiel
Minus	Minus	Formeln	$2 - 1 = 1$	Karl–Liebknecht–Straße
Viertelgeviertstrich	Hyphen	Bindestrich, Trennstrich	Karl-Liebknecht-Straße	5-7 Minuten
Halbgeviertstrich	n-dash	Bis-Strich, Gedankenstrich (Deutsch)	5–7 Minuten, plötzlich – ein gewaltiges Beben!	Karl–Liebknecht–Straße
Geviertstrich	m-dash	Gedankenstrich (Englisch)	Wow—What a long dash that is!	Cook the egg for 4—6 minutes

Tabelle B.2: Verschiedene Striche und ihr Einsatzzweck. Diese Tabelle ist so breit, dass sie rotiert dargestellt wird. Im PDF-Viewer sollte die Seite gekippt sein, sodass man den Kopf nicht drehen muss.

B.4 ABSTÄNDE UND ZEILENUMBRÜCHE

Das sind ganz normale Leerzeichen. Ein Zeilenumbruch wird als Leerzeichen interpretiert. Ich empfehle einen Satz pro Zeile, um kleinere Diffs im Versionskontrollsystem und bessere Übersichtlichkeit zu haben. Blocksatz ist Standard.

Zwei oder mehr Zeilen werden als neuer Absatz interpretiert. Man kann auch zwei Backslashes als Zeilenumbruch nehmen z.B. bei Tabellen, aber nicht im Fließtext, denn dann fehlt die Einrückung. Mit der Tilde kann man ein Leerzeichen vor Umbruch schützen, z.B. vor einer Zitierung wie dieser (Höffner u. a., 2017).

Es gibt auch halbe geschützte Leerzeichen, wie z. B. hier zwischen dem z und dem B, allerdings braucht man das nur sehr selten. Bei Zahlen und Maßeinheiten bietet es sich an, z.B. 50 000 kg statt 50 000 kg, aber dafür benutzt man lieber das Paket `siunitx`: 1 000 000 000 000 000, 50 000 kg.

B.5 ANFÜHRUNGSZEICHEN

Am besten mit dem `\enquote{}`-Befehl, dann werden automatisch die korrekten Anführungszeichen für die gewählte Sprache genommen: „Hier ein deutsches Zitat“. "Diese Anführungszeichen sind falsch", 'diese auch'.

GIT-TIPPS

Ein Schreckensszenario für alle Studenten ist es, Teile oder gar die ganze Abschlussarbeit zu verlieren, weil das Speichermedium kaputt geht oder die Dateien ausversehen gelöscht oder überschrieben werden. Nun weiß hoffentlich jeder theoretisch, dass Backups eine gute Idee sind aber die Frage ist, ob man das dann praktisch auch jede halbe Stunde macht und ob man dann hunderte Dateien wie „Bachelorarbeit-überarbeitet-v125-final-jetztwirklichfinal-needochnicht-aberjetzt.docx“ im Email-Posteingang hat und dann doch die falsche Version wiederherstellt. Daher gehe ich noch einen Schritt weiter und empfehle euch und Ihnen, die Abschlussarbeit in ein Git-Repository z. B. auf GitHub oder das Informatik GitLab zu stellen. Dies hat mehrere Vorteile:

1. hat man natürlich ein Backup, und das auf einem anderen System, und ohne seinen Rechner mit Kopien vollzumüllen.
2. kann man nicht nur den letzten Stand sondern *jeden* Commit wiederherstellen und sogar verschiedene Stände zusammenfügen.
3. sieht man dank der textbasierten LaTeX-Formate wie .tex und .bib in den Git-diffs genau, welche Änderungen wann durchgeführt wurden.
4. kann man mit Features wie Issues seine Arbeit organisieren und durch automatische Emails bei passenden Commits die Kommunikation mit den Betreuern vereinfachen.
5. erleichtert es das Arbeiten an verschiedenen Geräten, auch wenn es mal kein Internet gibt.

C.1 GITHUB UND GITLAB

Auf GitHub kann man sich einfach einen Account auf <https://github.com/> anlegen. Alternativ kann man als Informatikstudent auch einen Account auf <https://git.informatik.uni-leipzig.de/> anlegen, die grundlegenden Features sind ähnlich. Die browserbasierte LaTeX-Umgebung **OverLeaf** bietet auch eine Git-Integration.

Git-Tutorials gibt es z. B. unter <https://git-scm.com/docs/gittutorial> und <https://www.vogella.com/tutorials/Git/article.html>. Besonders tief muss man aber gar nicht in die Materie einsteigen, da man

ja nur alleine an einem Repository arbeitet und fast nie merges durchführen muss. Im Alltag benutzt man fast nur `git clone`, `status`, `pull`, `add`, `commit` und `push`.

Git ist allerdings für Textformate gedacht, die erzeugten PDFs sollten nicht mit im Repository stehen, da Binärdateien große Diffs erzeugen können.

ADVANCED PRINT TIPS (ENGLISH)

Warning: These are advanced L^AT_EX-tips. If they are too complicated for you, just ignore them, everything should be set up just fine! Just be sure to remind the print shop that you want *two-sided printing* „beidseitig drucken“ and it may be a few Euros more expensive.

D.1 ONE-SIDED VS TWO-SIDED

Save the environment, use two-sided printing! It also saves money and space in our cupboards! Two-sided printing also looks more professional in my opinion. In this ‘classithesis’ template, this is already done by including the ‘twoside’ argument to the ‘screpr’ document class in thesis.tex which mirrors the margins.

The title page and the first page of each chapter needs to be on the right side by adding empty pages if necessary, which is provided by the ‘openright’ option. View the PDF in two-sided mode and check if those pages are all on the right side and have odd page numbers. The title page is the first page in the document, so the PDF starts at the right side. However this is all already setup, so it should be good to go, just make sure to tell your printing shop that it is two sided because most students print on-sided and they often forget that you want it differently!

If you insist to use one-sided printing, you get the same margin on all pages. You can enable it by commenting out the twoside and openright options in thesis.tex:

```
\documentclass[toplevel, numbers=noenddot, %twoside, openright %a
    percent sign comments something out in LaTeX
    headinclude, footinclude, cleardoublepage=empty,
    abstract=on,
    BCOR=5mm, paper=a4, fontsize=11pt
]{screpr}
```

If you perform any changes here, check if the formatting in your document is still correct, as changing margins could move things around or destroy a table layout.

D.2 LATEX WARNINGS

Of course there should never be any errors when compiling a L^AT_EX document but what about all those hundreds of warnings that you collected and ignored over the months? The log can have a thousand lines or more because pdf_latex is very verbose but some nasty surpri-

ses can be hidden there so I recommend you to go through it at least once, even though you cannot fix all of them. You can remove some of the clutter in vim via `:%s/^.*\usr\share\:q\texmf-dist\/*$\n//` (adapt to your LaTeX installation path).

D.2.1 *Overful Hboxes*

Overful hboxes are elements going over your right margin, which should not happen. Overful and underful hboxes are included in your log but you can visualize overful hboxes using `\setlength{\overfullrule}{20pt}` (don't forget to take this out afterwards). Sometimes an overful hbox is reported but not actually visible because only the invisible part of an element, such as a table, runs over the margin, in which case you can ignore it.

D.3 COLOR

Colored pages are about ten times more expensive than grayscale, so you need to find them, reduce the number of them as much as possible and then tell the printing studio how many there are for calculating the price.

D.3.1 *Colored Text Elements*

At first I had a huge number of colored pages because all my links and listings were colored, which looks great on paper but is very costly. However you can still keep a screen version with all the nice colors intact. Bonus points if you can implement a switch between a print and a screen version. I didn't have the time left for that and you probably don't and shouldn't, but if you do, please tell me!

```
\hypersetup{%
colorlinks=false, % print version
%colorlinks=true, % screen version
[...]
pdfborder={0 0 0}, % I don't like link borders but if you do
    those are perfect for that purpose as they are only shown on
    screen, not in print! Requires colorlinks=false to be shown.

\definecolor{listkeyword}{gray}{0.3}
\lstset{
%keywordstyle=\color{cyan}, % screen version
keywordstyle=\color{listkeyword}, % print version
[...]
}
```

D.3.2 Colorcheck script

Ghostscript can detect all pages that contain color [Source: StackOverflow](#). The colorcheck script will tell you, how many colored and grayscale pages your document has and show you the pages with color on it and the average values for the CMYK channels in that order. The 'tail -n +6' is adapted to Ghostscript 9.52 and may break in the future and removes the ghostscript version and license information.

```
gs -o - -sDEVICE=inkcov $1 | tail -n +6 | sed ':a;N;$!ba;s/\n
  / /g' > /tmp/colorlog # each page on one line
echo -n "Black and white pages: "
grep -e "0.00000 0.00000 0.00000 [01]." /tmp/colorlog | wc
-l
echo -n "Colored pages: "
grep -v -e "0.00000 0.00000 0.00000 [01]." /tmp/colorlog |
wc -l
grep -v -e "0.00000 0.00000 0.00000 [01]." /tmp/colorlog
```

Example:

```
\$ colorcheck main.pdf
Black and white pages: 128
Colored pages: 7
Page 1 0.00000 0.00264 0.00264 0.01284 CMYK OK
Page 21 0.03851 0.03842 0.03844 0.03065 CMYK OK
Page 22 0.07243 0.06588 0.07243 0.10790 CMYK OK
Page 66 0.01619 0.01619 0.01399 0.05655 CMYK OK
Page 81 0.09786 0.11218 0.10922 0.02548 CMYK OK
Page 82 0.15376 0.15006 0.14548 0.06880 CMYK OK
Page 84 0.04707 0.04707 0.03834 0.03834 CMYK OK
```

D.3.3 CMYK Color Space

You only need to know three things about CMYK:

1. it stands for Cyan, Magenta, Yellow and Key (black)
2. black pages need to only use the key channel
3. not all RGB values can be represented in CMYK, such as a bright red (255,0,0)

A page that gets reported by colorcheck may contain a PDF picture that looks perfectly grayscale to the human eye but that has its black value mixed from the other colors, not the black channel. If it is a picture that you didn't produce yourself from within LaTeX, you can convert it using Ghostscript:

```
gs \
-sOutputFile=output.pdf \
-sDEVICE=pdfwrite \
```

```

-sColorConversionStrategy=Gray \
-dProcessColorModel=/DeviceGray \
-dCompatibilityLevel=1.4 \
-dNOPAUSE \
-dBATCH \
input.pdf

```

To ensure better looking colors, prefer CMYK primary colors to RGB. The following lines are an excerpt from a Tikz picture of a Venn diagram:

```

% CMYK
\node[venn circle = magenta] (SW) at (0,0) {};
\node[venn circle = cyan] (QA) at (60:0.17\textwidth) {};
\node[venn circle = yellow] (QB) at (0:0.17\textwidth) {};
+% RGB
+% \node[venn circle = red] (SW) at (0,0) {};
+% \node[venn circle = blue] (QA) at (60:0.17\textwidth) {};
+% \node[venn circle = green] (QB) at (0:0.17\textwidth) {};

```

D.4 PRINTING

I went to **REPRO Graf** in Leipzig, who printed and bound it for me quickly and for a good price. I didn't choose any fancy options, just standard A4 80g paper and hardcover binding. The only problem was that they printed it one-sided as that is their default, however I just kept it as a personal copy. Because two-sided margins differ between even and odd pages, you cannot submit a one-sided printing of a two-sided document, so communicate this very clearly in advance!

ERKLÄRUNG

Ich versichere, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt habe, insbesondere sind wörtliche oder sinngemäße Zitate als solche gekennzeichnet.

Mir ist bekannt, dass Zuwiderhandlung auch nachträglich zur Aberkennung des Abschlusses führen kann.

Ich versichere, dass das elektronische Exemplar mit den gedruckten Exemplaren übereinstimmt.

Leipzig, 31.09.2023

Paul Keller