

Prüfungsleistung im Modul Verarbeitung von Textdaten

M.Sc. Informatik

Nils Reiter

20. November 2023

In diesem Dokument finden Sie Informationen zur Modulprüfung im Modul „Verarbeitung von Textdaten“, die aus der Durchführung einer praktischen Arbeit und dem Verfassen eines wissenschaftlichen Textes dazubesteht. Im Modulhandbuch sind dafür etwa 180 Stunden Zeitaufwand vorgesehen, das sind etwa **viereinhalb Wochen in Vollzeit!** Planen Sie diese Zeit auch ein.

1 Themen

Grundlage für Themen können grundsätzlich Inhalte aus allen Veranstaltungen aus dem Modul (oder dem restlichen Studium) sein, solange die unten genannten Voraussetzungen erfüllt sind. Sie sollten, schon im eigenen Interesse, ein Thema wählen mit dem Sie sich schonmal beschäftigt haben (und das Sie interessant finden).

- Jedes Thema *muss* einen praktischen Anteil beinhalten. Dies kann z.B. eine Korpusuntersuchung sein, eine Implementierung eines (kleinen) Experimentes, eine systematische Evaluation mit Versuchspersonen, die Anwendung existierender Tools auf neue Daten oder etwas anderes.
- Themen *müssen* vorher abgesprochen werden. Das geht am einfachsten per Mail, indem Sie formlos eine kurze Beschreibung (3-4 Absätze) des Themas per E-Mail an nils.reiter@uni-koeln.de schicken. Falls nötig, treffen wir uns dann in meiner Sprechstunde, aber ich hoffe wir können das schriftlich regeln.

2 Hausarbeit (formale Vorgaben)

Die Hausarbeit sollte maximal acht Seiten lang sein.¹ Verwenden Sie eine Schriftart *mit Serifen* (z.B. Times New Roman) und einen einfachen Zeilenabstand mit einer Schriftgröße von 11 pt. Als Seitenränder stellen Sie links 2,5cm und sonst 2cm ein.

¹Das wird nicht reichen, um alles aufzuschreiben. Sich auf die wesentlichen Dinge zu konzentrieren gehört zu den Aufgaben beim wissenschaftlichen Schreiben.



Abbildung 1: Ein Hund

- Die Hausarbeit muss per E-Mail an nils.reiter@uni-koeln.de abgegeben werden. Der Text der Arbeit sollte eine einzelne PDF-Datei sein. Weitere Dateien (Code, Ressourcen etc.) bitte als zip-Datei. Wenn Sie nicht innerhalb von ein paar Tagen eine Bestätigung bekommen haben, zögern Sie nicht nachzufragen.
- Bei unvorhersehbaren Hindernissen (z.B. Krankheit, Rechner kaputt, ...) kann die Abgabefrist verlängert werden (ggf. ist ein Nachweis, also Attest o.ä. erforderlich). In dem Fall bitte umgehend bei mir melden.
- Vor dem eigentlichen Text stehen Deckblatt und Inhaltsverzeichnis, nach dem eigentlichen Text Literaturverzeichnis.
- Die Seiten und auch die Hauptabschnitte sind nummeriert (so wie in diesem Dokument).

Gliederung In der Hausarbeit sollten die folgenden Bereiche abgedeckt sein:

- Einleitung/Motivation: Was ist das Thema, warum beschäftigt man sich damit, warum ist es interessant?
- Forschungsstand: Was ist der Forschungsstand für das Thema? Welche anderen Arbeiten gibt es? Welche Ressourcen (z.B. Korpora) gibt es? Hier sind knapp die wesentlichen Gesichtspunkte wiederzugeben.
- Bericht zum praktischen Teil: Was genau wurde gemacht, und wie? Was sind die Ergebnisse davon? Welche Schlussfolgerungen kann man daraus ziehen?
- Fazit/Konklusion: Was haben wir gelernt? Was haben Sie gelernt? Was wären Schritte die „man“ als nächstes tun könnte/müsste/sollte?

Die Hausarbeit ist *keine* vollständige Projektdokumentation, d. h. es wird Teile oder Module geben, die in der Arbeit nicht vorkommen. Der Text sollte gleichzeitig für Außenstehende (Fachleute) nachvollziehbar sein.

Referenzen / Zitate Die Arbeit sollte der „guten wissenschaftlichen Praxis“ entsprechen. D.h. insbesondere, dass nicht-eigene Inhalte gekennzeichnet werden müssen. Das gilt auch für indirekte und nicht-wörtliche Zitate. Wie genau Sie zitieren sei Ihnen überlassen, aber es sollte *einheitlich* sein.

Beliebt in vielen Fächern ist das sog. „author-year“-Schema, also z.B.:

Müller (2019) stellt fest dass, Ferien gut sind. Damit widerspricht sie früheren Ergebnissen (Schneider, 2017; Smith et al., 2018), die das Gegenteil herausgefunden haben.

In jedem Fall bestehen die Referenzen aus zwei Teilen: Innerhalb des Textes werden die Übernahmen gekennzeichnet, und am Ende findet sich ein Literaturverzeichnis, in dem alle verwendeten Quellen aufgelistet werden.

Tipp: Mit einer Literaturverwaltungssoftware lässt sich vieles davon automatisieren. Mögliche Startpunkte: Zotero, Citavi, Endnote, BibTeX.

Abbildungen, Tabellen, Listings Formeln, Abbildungen, Tabellen und Listings mit Code- oder Pseudocode-Snippets werden durchnummeriert, so dass im Text mit Nummer darauf referiert werden kann. Abbildungen, Tabellen und Listings bekommen außerdem eine Beschriftung („caption“), wie in Abbildung 1 gezeigt. Setzen Sie diese Dinge bitte *nicht* gesammelt ans Ende, sondern innerhalb des Textes in der Nähe der Stelle an der sie erwähnt werden. Verweise innerhalb des Textes sehen dann z. B. so aus:

Abbildung 5 zeigt dass, ...

Wie in Formel 2 ersichtlich, ...

Die Ergebnisse (Tabelle 4) zeigen, dass ...

Code oder pseudocode bitte in einem **monospace**-Font setzen. Vollständige Programme bitte *nicht* in den Anhang des Dokumentes setzen, sondern ggf. als digitalen Zusatz (zip-Datei) in kommentierter Form mit abgeben.

Englischsprachige Begriffe Für viele englischsprachigen Begriffe gibt es keine passenden deutschsprachigen. In dem Fall lieber den englischen verwenden, als eine deutsche Übersetzung zu erfinden. Es gibt verschiedene Möglichkeiten wie englische Begriffe formatiert werden können. Am wichtigsten ist dass Sie das konsistent machen. Als Vorschlag die Regeln die ich verwende:

- Englische Begriffe werden klein und kursiv geschrieben: „Die *accuracy* beträgt 75 %“
- Verbindungen aus deutschen und englischen Wörtern werden mit Bindestrichen verbunden: „Der *accuracy*-Wert alleine ist aber nicht aussagekräftig, denn ...“

3 Tipps

- Verwenden Sie eine Rechtschreibprüfung.
- Suchen Sie sich Testleser*innen. Das können Freundinnen und Mitbewohner sein, aber auch Kleingruppen, in denen Sie die Texte gegenseitig gegenlesen.
- Achten Sie auf den roten Faden: Es sollte in ein durchgehender Text sein, in dem die einzelnen Teile logisch zusammenhängen.
- Sprache und Stil: Achten Sie auf einen präzisen, sachlichen und verständlichen Schreibstil. Der von Ihnen produzierte Text muss nicht unterhalten oder originell sein, und auch einen Spannungsbogen müssen Sie nicht erfinden. Die Hausarbeit ist *kein* Essay oder ein Meinungstext, sondern ein wissenschaftlicher Bericht über Experimente.
- Gehen Sie schrittweise vor: Erstes Ziel ist eine *inhaltlich vollständige Fassung*. Diese wird dann in ggf. mehreren Runden gekürzt, verdichtet und auf den Punkt gebracht. Dabei ringt man um einzelne Worte.
- Je länger man an einem Text arbeitet, desto besser wird er.

LaTeX

- Mit der Klasse `scrartcl` machen Sie alles richtig.
- Mit `\usepackage[ngerman]{babel}` bekommen Sie eine deutsche Silbentrennung.
- Nützliche Pakete: `listings`, `booktabs`
- Inhaltsverzeichnisse sollten Sie automatisch erzeugen (lassen).
- Machen Sie sich mit `bibtex` oder `biblatex` vertraut, um ein Literaturverzeichnis automatisch einheitlich erzeugen zu lassen.

Word/OpenOffice

- Benutzen Sie Formatvorlagen, um die Formatierung einheitlich zu machen.
- Inhaltsverzeichnisse sollten Sie automatisch erzeugen (lassen).
- Zur Erzeugung von Literaturverzeichnissen gibt es Software: `zotero` ist kostenlos und open source.

3.1 Häufige Probleme

Häufige Probleme sprachlicher und stilistischer Art sind:

- Unklare Bezüge: Überprüfen Sie vor der Abgabe, ob der Text eindeutig ist. Insbesondere bei Anaphern („dies“, „dort“, „es“, ...) ist ggf. Vorsicht geboten.
- Falsch verwendete Begriffe: Wichtigstes Kriterium für wissenschaftliches Schreiben ist Präzision. Fachbegriffe dienen dazu, dies zu ermöglichen. Verwenden Sie Begriffe sorgfältig. Teilweise weicht die Bedeutung wissenschaftlicher Begriffe vom Alltagsgebrauch ab. Wenn Sie Begriffe selbst definieren, tun Sie dies explizit.
- Falsche Kausalzusammenhänge oder Gegensätze: Achten Sie darauf, welche Zusammenhänge sie mit „da“, „also“, „deshalb“, „dennoch“ etc. herstellen.

4 Häufig gestellte Fragen (FAQ) – mit Antworten!

4.1 Kann ich die Arbeit testweise abgeben und ggf. nochmal überarbeiten?

Nein, bewertet werden nur endgültig abgegebene Arbeiten.

4.2 Muss ich X erklären?

Wenn X für das Verständnis der Arbeit wichtig ist, muss es erklärt werden.

4.3 Wie zitiere ich Internetquellen?

Versuchen Sie zunächst herauszufinden, ob es zu der Quelle eine wissenschaftliche Referenz gibt. Oft basieren Blog-Einträge von Wissenschaftler*innen auf Veröffentlichungen – dann besser die zitieren. Wenn es für eine Internetquelle keine andere Grundlage gibt, zitieren Sie sie wie die anderen auch mit Name und Jahr im Text. Im Literaturverzeichnis sollte dann in jedem Fall der Link und ggf. ein Titel auftauchen. Verwenden Sie als Jahr idealerweise das Jahr, in dem der Inhalt erstellt wurde. Wenn möglich, verwenden Sie einen permanenten Link, also einen der sich nicht ständig ändert.

4.4 Wie zitiere ich Wikipedia?

Grundsätzlich so wie andere Internetquellen auch. Bei Wikipedia sollten Sie aber sicherstellen, dass sie auf eine spezifische *Version des Artikels* verlinken. Im Übrigen finden Sie auf Wikipedia-Artikeln üblicherweise eine Liste mit wissenschaftlichen Referenzen, auf denen der Wikipedia-Artikel seinerseits basiert. Ggf. ist es sinnvoll, direkt diese Quellen zu zitieren.

4.5 Ich habe keine Idee für ein Thema

Wenn Ihnen nichts direkt einfällt, hilft vielleicht die Materialien aus den Veranstaltungen nochmal anzuschauen, die Sie besucht haben. Ein paar Ansatzpunkte, die man meistens machen kann: Lässt sich die Arbeit X auf eine andere Sprache oder andere Textsorte übertragen? Für welchen Anwenderkreis macht eine graphische Oberfläche für welches Problem Sinn, die es noch nicht gibt? Gibt es verwandte Phänomene, die sich ähnlich bearbeiten lassen? Sind Abwandlungen/Vereinfachungen einer Arbeit möglich, und wie gut wären die? Natürlich können Sie sich auch an mich wenden, wenn das auch nicht hilft.

4.6 Kann ich die Hausarbeit auch in englischer Sprache schreiben bzw. die Präsentation in englischer Sprache halten?

Yes.