



Wissenschaftlich Schreiben

Ulf Leser

Top-7 Fehler

- Ansicht, dass wissenschaftliche Texte kompliziert sein müssen, lange geschachtelte Sätze haben, komplizierte Spannungsbögen und möglichst viele Fremdwörter
- Logische Fehler in Sätzen
- Bilder/Tabellen, die nirgends referenziert werden
- Unvollständige bibliographische Angaben
- Fehlendes Abstract oder fehlendes Inhaltsverzeichnis
- Umgangssprachlicher, aufgepeppter Sprachstil
 - „sonst klingt es so langweilig“
- Erratisch gesetzte oder völlig fehlende Absätze / Sections

Vorwort

- Wissenschaftliches Schreiben ist anders als literarisches
 - Fokus auf **Verständlichkeit** (statt Spannung, Betroffenheit, ...)
 - Regt nicht zum Nachdenken an, sondern **präsentiert Ergebnisse**
 - Aussagen müssen belegt sein – durch Text oder durch Referenzen
 - Sollte **einfach** zu lesen sein – „querlesen“ ermöglichen
 - Hat einen ziemlich standardisierten Aufbau
- Aber wissenschaftliches Schreiben ist auch eine Kunst
- Kunst lernt man durch **Übung** (und ein bisschen Begabung)
- Lesen Sie anderer Arbeiten kritisch – man lernt viel

Inhalt

- Inhalt
- Aufbau
- Schreibstil
- Formatierung
- Bibliographie
- Zitieren / Plagiate
- Weitere Elemente
- Letzte Worte

Inhalt Seminararbeit

- Motivation des Themas
 - Mit möglichst **konkreten Beispielen**
- Formale Problemstellung
 - „Thema“ wird zum **klar definierten Problem**
- Stand der Technik
 - Wie hat man das Problem bisher gelöst, was war daran **unbefriedigend**
- Neuer Ansatz (das eigentliche „Thema“)
 - Wie funktionieren sie, was ist dadurch **besser**
- Empirische Evaluation
 - Quantitativer Vergleich mit anderen Verfahren
- Diskussion
 - Was wurde in der Evaluation nicht gezeigt, welches weitere **Verbesserungspotential** gibt es, welche **Problemvarianten** gibt es, etc.?
- Selbstständigkeitserklärung mit Unterschrift

Aufbau

- Titelseite
- Abstract / Zusammenfassung (ca. 20 Zeilen)
- Inhaltsverzeichnis
- 3-5 Kapitel
 - Grundlagen und verwandte Arbeiten
 - Formale Problemstellung
 - Vorgestellte Lösung
 - Evaluation
- Diskussion und Zusammenfassung
- Bibliographie
- **Anhang** (häufig als Möglichkeit vergessen)
- Selbstständigkeitserklärung mit Unterschrift

Inhaltsverzeichnis

- **Standard:** Abstract, Einleitung, Grundlagen und verwandte Arbeiten, Problemstellung, Lösung, Evaluation, Diskussion und Zusammenfassung
 - Wenn möglich, sollte man sich ungefähr **daran halten**
- **Format**
 - IV auf eigener Seite
 - Kapitel und Sections reichen – nicht tiefer gehen (bei 10 Seiten)
 - Verschiedene Layouts sind möglich – lesbar, zurückhaltend
 - Nach Möglichkeit auf eine Seite passend (bei 10 Seiten)
 - **Seitennummer** (rechtsbündig)
 - Etwa **gleiche Zahl Unterelemente** pro Element sind angenehm
 - Keine Elemente mit nur einem Unterelement (mergen)
 - Kurze, **klare Überschriften**

Wichtige Bestandteile

- Abstract (200-300 Wörter)
 - Kurze Motivation, Neuheit des Ergebnisses, Ausblick auf die konkreten Inhalte der Arbeit
 - Ziel: Interesse an der Arbeit wecken
- Einleitung (1-2 Seiten)
 - Längere Motivation des Themas
 - Stand-der-Technik andeuten und **Mängel** darstellen
 - Übersicht über im Paper vorgestellte Methode und deren Vorteile
 - Übersicht über Kapitel der Arbeit
 - „In Kapitel zwei werden wir ... In Kapitel drei wird ...“
- Zusammenfassung und Diskussion
 - Ein Absatz, der die wichtigsten Ergebnisse zusammenfasst
 - 1-3 Absätze, die die Ergebnisse kritisch diskutieren

Entstehungsprozess: Top-Down / Bottom-Up

- Mein Tipp: **Verfeinern Sie zunehmen**
 - Kapitel und Sections, dann Subsections
 - Pro Subsection Stichpunkte, die den **roten Faden** wiedergeben und die Reihenfolge der Argumente festlegen
 - Was will ich in dieser Subsection sagen?
 - Alle **Übergänge prüfen**
 - Sind Subsections in richtiger Reihenfolge? Gibt es exzessiv viele Referenzen nach vorne (wie in Kapitel 2 gezeigt, ...) oder nach hinten?
 - Vollständigkeit prüfen
 - Habe ich alles gesagt, was ich in den Exzerpten zur Literatur für wichtig hielt?
 - Ausformulieren
 - **Überarbeiten**
 - **Nochmal überarbeiten**

Schreibblockaden

- Bottom-Up kann Blockaden lindern – Textbausteine produzieren, noch nicht auf Fluss achten
 - Top-Down ist für disziplinierte Schreiber
- Erste Fassungen sind nie perfekt - **erstmal schreiben**, später überarbeiten

Inhalt

- Inhalt
- Aufbau
- **Schreibstil**
- Formatierung
- Bibliographie
- Zitieren / Plagiate
- Weitere Elemente
- Letzte Worte

Schreibstil [Dies ist kein „Scientific Writing“ Kurs]

- Neutral und zurückhaltend
- Kurze Sätze, keine komplizierten Konstruktionen
- „Jedes Adjektiv muss genehmigt werden“ [Die Zeit, ?]
- Aktiv statt passiv, Verben statt Substantive (vor allem in E)
 - „Eine Demonstration der Überlegenheit des XYZ-Algorithmus gegenüber dem ABC –Algorithmus in Bezug auf Worst-Case Komplexität wurde in [1] vorgenommen.“
 - „Meier et al. zeigen in [1], dass der XYZ-Algorithmus mit $O(\log(n))$ eine bessere Worst-Case Komplexität als ABC hat.“
- Keine Superlative, keine Umgangssprache
- „wir“ statt „ich“

Schreibstil 2

- **Fachtermini** (korrekt) verwenden
 - Bei mir auch ruhig die Englischen
- Füllwörter **alle** entfernen (keine Gnade!)
 - Dabei, eigentlich, genauer gesagt, selbstverständlich, ...
- Subsections maximal eine Seite
 - Sonst weiter untergliedern
- Beispiele, Abbildungen, Tabellen, **Aufzählungen**, Pseudocode verwenden
 - Auch im Textfluss: „Dies hat die folgenden Vorteile: (1) Es ist schneller, (2) es ist besser wartbar, und (3) es ist robuster“
- **Absätze gliedern** ihren Text logisch
- Einheitliches Tempus

Formatierung

- Wie bei Folien – **Zurückhaltung** ist Trumpf
- Nur ein Font (z.B.: Times oder Arial)
- Nur wenig verschiedene Schriftgrößen
 - Kapitel: 14 pt, fett, 24 vor, 12 nach, nummeriert
 - Section: 12 pt fett, 12 vor, 6 nach, nummeriert
 - Subsection: 12 pt fett, 12 vor, 6 nach, nicht nummeriert
 - Text: 10 pt, nach jedem Absatz 3pt
- Zeilenabstand 1,5 Zeilen (korrekturfreundlich)
- Verwenden Sie LaTeX oder Absatzvorlagen (Word)
- Header: Titel abgekürzt; Footer: Seitennummer
- Bei Abschlussarbeiten: Kapitel auf neuen Seiten
- **Aufzählungen** (Bullets) sind wunderbar!

Formatierung 2

- Am besten im Text weder **fett** noch *kursiv* noch unterstreichen
 - Aber wenn, dann einheitlich
- Bei Word – Formeleditor verwenden
 - Nicht optimal, aber auch nicht so schlecht
- **Absätze im Blocksatz**, um die Struktur zu betonen
 - Silbentrennung nicht vergessen

Deckblatt

- Titel
- Autor
- Datum
- Art der Arbeit
- Veranstaltung
- Betreuer

HUMBOLDT-UNIVERSITÄT ZU BERLIN
MATHEMATISCH-NATURWISSENSCHAFTLICHE FAKULTÄT
INSTITUT FÜR INFORMATIK

Flow Maximization in Networks

Proseminar

Danielle Melis Kochisarilar

12.6.2017

Dozent Prof. Dr. Ulf Leser

Humboldt-Universität zu Berlin
Institut für Informatik

Suffix Trees und Suffix Arrays

Nele Harder, 565634
harderne@informatik.hu-berlin.de
Monobachelor Informatik

Betreuer: Prof. U. Leser
Proseminar Wissenschaftliches Arbeiten
Sommersemester 2017

15.06.2017

Inhalt

- Inhalt
- Schreibstil
- Formatierung
- **Bibliographie**
- Zitieren / Plagiate
- Weitere Elemente
- Letzte Worte

Drei goldene Regeln [The12]

- Geben Sie zu **allen Informationen aus Arbeiten anderer** diese Quellen an
 - Alle wichtigen Aussagen, die Sie nicht selber beweisen / untermauern, sollten durch Quellen belegt werden
 - Alle Gedanken, die Sie von anderen übernehmen, müssen durch Quellen belegt werden
- Positionieren Sie die Quellverweise in Ihrem Text eindeutig
 - Es muss klar sein, wo der übernommene Gedanke / Aussage anfängt und wo er aufhört
- Quellangaben müssen so gestaltet sein, dass **die Quelle möglichst einfach auffindbar ist**
 - Vollständige bibliographische Angaben, aber auch nicht mehr

Bibliographie

- Bibliographie: Liste **aller referenzierten Arbeiten (und nur dieser)** am Ende der Arbeit
- Bibliographische Angaben müssen stets vollständig sein
- Sinn
 - **Nachprüfbare Belege** zu Aussagen im Text
 - Anerkennung der Leistung anderer
 - Klare Trennung eigener Gedanken von Fremden
- Unterscheidung
 - Zitation / Kurzbeleg: Verweis aus dem Text auf die Referenz
 - Referenz: Eintrag in der Bibliographie

Bibliographische Angaben

- Buch: Author, title, year, Verlag
- Journals: Author, title, year, journal, volume, (issue), pages
- Conference: Author, title, year, conference name, location
- Reports: Author, title, year, type of work, institution, number
- Buchkapitel (editiert): Author, title, year, Buch, Verlag, Editoren
- Beispiele
 - Rawald, T., Sips, M., Marwan, N. and Leser, U. (2015). "Massively Parallel Analysis of Similarity Matrices on Heterogeneous Hardware". Int. Workshop on Data (Co-)Processing on Heterogeneous Hardware Brussels, Belgium
 - Rheinländer, A., Heise, A., Hueske, F., Leser, U. and Naumann, F. (2013). "SOFA: An Extensible Logical Optimizer for UDF-heavy Dataflows". CoRR/abs:1311.6335.
 - Rheinländer, A., Heise, Hueske, Leser, and Naumann, (2015). "SOFA: An Extensible Logical Optimizer for UDF-heavy Data Flows " Information Systems 52: 96 - 125.
 - Hakenberg, J., Plake, C. and Leser, U. (2010). Ali Baba: A Text Mining Tool for Complex Biological Systems. In Lodhi, H. and Muggleton, S. (ed): Elements of Computational Systems Biology, Wiley & Sons

Zitation (Kurzbeleg) - Verschiedene Stile

- **Drei Autorennachnamen Anfangsbuchstaben**, dann Jahr
 - [RHH+15] Rheinländer, A., Heise, A., Hueske, F., Leser, U. and Naumann, F. (2015). "SOFA: An Extensible Logical Optimizer for UDF-heavy Data Flows " Information Systems 52: 96 - 125.
 - [Sea05] Searls, D. B. (2005). "Data Integration: Challenges for Drug Discovery." Nature Reviews Genetics 4: 45-58.
- **Nummerieren**
 - Sortierung: Alphabetisch nach Erstautor oder nach Reihenfolge der Erstzitation im Text
 - [1] Searls, D. B. (2005). "Data Integration: Challenges for Drug Discovery." Nature Reviews Genetics 4: 45-58.
- **Autorennamen**
 - Dann lässt man die Kurzbelege in der Bibliographie weg
 - (Rheinländer et al. 2015), (Rheinländer und Leser, 2015)

Feinheiten

- Immer **alle Autorennamen**. Alternativ: Ab 5/10 „et al.“
 - Stark, A., Lin, M. F., Kheradpour, P., Pedersen, J. S., Parts, L. et al. (2007). "Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures." Nature 450: 219-32
- Vornamen abkürzen oder ausschreiben
 - Aber: Einheitlich!
- Manche ...
 - setzen Titel oder Journal kursiv – muss man nicht
 - setzen Volume+Seiten fett – muss man nicht
 - lassen Titel weg (nur Autoren) – soll man nicht
- Mehrere Zitationen: **In einer Klammer**
 - „Wie in [3,5,6] gezeigt, ...
 - „Wie in [AKF+13, BH12, Bel10] gezeigt, ...“

Weitere Feinheiten

- Bei Zitationsstil [1] und [RVS+15] keine weiteren Klammern
 - Nicht: „Wie in ([1]) gezeigt, ...“
- Kurzbelege kommen in den Satz, vor den finalen „.“
 - „Dies ist ein effizienter Algorithmus [4]. Dagegen ...“
- Guter Stil: Öfter mal **Autorenname explizit benennen**
 - „Searls hat gezeigt, dass $A \neq B$ [Sea95].“
 - „Rheinländer und Leser beschreiben den PieJoin Algorithmus [RL13], der die Eigenschaft hat, dass ...“
 - „Rheinländer et al. beschreiben den PieJoin Algorithmus [RLR+13], der die Eigenschaft hat, dass ...“

Zitieren (und Plagiate)

- Jede **wörtliche Übernahme** mit nicht trivialer Länge muss als Zitat gekennzeichnet werden
 - In Hochkomma, mit Zitation
- Wörtliches Zitieren ist in der Informatik sehr unüblich
 - Wir haben wenig Personenkult
 - Ganz wenige Ausnahmen: Weizenbaum, Turing, Gates, ...
- Sehr oft übernimmt man Gedankengänge, Aussagen, ...
 - Zitation am Ende eines Absatzes, Grenzen des belegten müssen erkennbar sein
 - **Eigene Worte** verwenden

Beispiel (Related Work Sections aufmerksam lesen!)

In the following we review the most important approaches for evaluating set containment joins, namely signature-, hash-, and tree-based algorithms and joins based on inverted indexes.

Signatures, also called superimposed coding [17], are a method to compactly represent sets using bit vectors. Instead of reserving one bit for every item of the domain, these algorithms use a much smaller vector, hashing each item of the domain to k bits in this vector. A set is the bitwise-OR of the vectors of all its items. This enables to compare two sets very efficiently using bitwise operations. However, there is also a downside: this technique introduces false positives, which have to be filtered out in a second step. The main difference between signature-based algorithms is the way they store the signatures.

Signature nested-loop and sequential signature le joins [6, 11] write them sequentially to a le, while signature tree joins utilize a tree structure [4]. Hash-based joins partition the tuples into different buckets using an appropriate hash function. The tuples found in corresponding buckets are then joined using one of the other techniques. The most prominent algorithms in this area are partitioned set join [16], adaptive pick and sweep join, divide-and-conquer set join, and adaptive divide-and-conquer set join [14, 15]. There is also some overlap with signature-based techniques. For instance, signature hash, extendable signature hashing, and lattice set join [5, 6, 13] use signatures for hashing the tuples into buckets.

Inverted indexes, also called inverted les, are an access method commonly used in ...

Plagiate

- Sind komplett **inakzeptabel**
 - Diebstahl von Gedankengut / Ideen
 - Andere Gedanken am besten als eine Art Patent betrachten
 - Ausschluss von Seminar, Abschlussarbeiten sofort nicht bestanden, auch **nachträgliche Aberkennungen von Titeln** und strafrechtliche Konsequenzen möglich
- Wörtliche Übernahme **ohne Kennzeichnung** ist ein Plagiat
- Großflächige Darstellung fremder Gedanken **mit korrekter Kennzeichnung** ist kein Plagiat
 - Aber auch keine originäre wissenschaftliche Leistung
 - Typisch für Übersichtsarbeiten, Seminararbeiten, Related Work
 - Man kann z.B. in der Intro schon angeben: „Die folgenden Gedanken wurden aus [X] übernommen, wenn nicht anders gekennzeichnet“

Selbstplagiate

- Sehr **unterschiedliche Einschätzungen**
- Beispiel
 - Workshop Paper betrachten manche nicht als Veröffentlichung – darf (ggf. in erweiterter Form) erneut publiziert werden
 - Journal Artikel (>15 Seiten) als **erweiterte Fassungen** von Konferenzbeiträgen (8-12 Seiten)
 - Daumenregel: >30% neuer Inhalt (viele Diskussionen)
 - Reports gelten nicht als (peer-reviewed) Veröffentlichung
 - Technical Reports „as is“ in Journalen veröffentlichen
 - TR durchlaufen keinen Review-Prozess – **Schnelle Veröffentlichung**
 - TR bleibt Open Access, Journalartikel nicht
 - **Kapitel aus Dissertationen** sehr oft eigene Veröffentlichungen
 - Früher: Nachher publizieren; Heute: Vorher publizieren
 - Kumulative Promotionen

Inhalt

- Inhalt
- Schreibstil
- Formatierung
- Bibliographie
- Zitieren / Plagiate
- **Weitere Elemente**
 - Verzeichnisse, Fussnoten, Bilder, Tabellen
- Letzte Worte

Verzeichnisse

- Immer: **Inhaltsverzeichnis und Bibliographie**
 - Seminararbeit: Sonst nichts
- Texte ab 30 Seiten Länge: Vielleicht Tabellen- /
Abbildungsverzeichnis
- Dissertationen: Ggf. Abkürzungsverzeichnis, Glossar
- Lieber weniger als mehr

Fußnoten

- Sinn: Nicht für alle Leser relevante Information, die den **Textfluss stören** würde
- Informatiker verwenden sie **sehr sparsam**
 - Unwichtiger Inhalt – weglassen
 - Wichtiger Inhalt – in den Text einfügen
- Nicht (!)
 - **für Referenzen** verwenden
 - zur Darstellung **anderer Meinungen** verwenden
 - ¹ Im Unterschied dazu meinen Starlinger und Müller [SM14], ...
 - zur „echten“ Erklärung verwenden
 - ¹ Genauer gesagt gilt, dass
- Am besten: **Ohne auskommen**
 - Ausnahme: URLs

URLs - Zwei Stile möglich

- In die Bibliographie, mit eigener Zitation
 - Zitation braucht ggf. Phantasie [App15], [Wp13a]
 - [App16]: <http://www.apple.de/latestnews>, visited 13.4.2016
 - [Wp16a]: <http://www.wikipedia.com/topic1>, visited 14.5.2016
 - [Wp16b] : <http://www.wikipedia.com/topic2>, visited 14.5.2016
- Als Fußnote
 - Mein Favorit – keine Zitation, klare Trennung wissenschaftlicher Referenzen von nicht wissenschaftlichen
- Immer mit angeben: Datum des letzten Besuchs

Bilder / Tabellen / Algorithmen

- Werden auf der Seite horizontal **zentriert**
- Werden nicht von Text umflossen
- Haben eine zentrierte, links- und rechts eingerückte, im Font leicht verkleinerte **Unterschrift** (Caption)
 - Bild+Caption müssen selbsterklärend sein
- Werden **fortlaufend nummeriert**
 - Ab 100 Seiten mit Kapitelnummer als Präfix
 - Getrennte Nummern für Abbildung, Tabelle, Algorithmus
- Werden im Text **referenziert**
 - Sonst sieht ein Profi-Leser sie nicht an

Können kombiniert werden

- Zum Platz sparen oder ...
 - Eher nicht bei Seminar- / Abschlussarbeiten
- Wegen des inhaltlichen Zusammenhangs

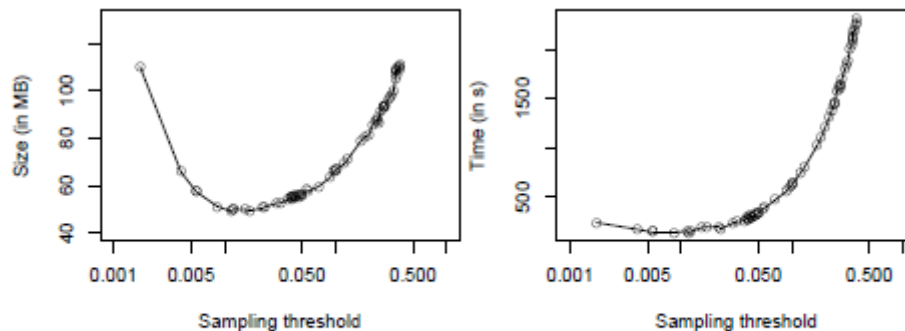


Figure 4: Impact of iTong's sampling threshold on index size (left) and indexing time (right) for HEL with 2,560 strings.

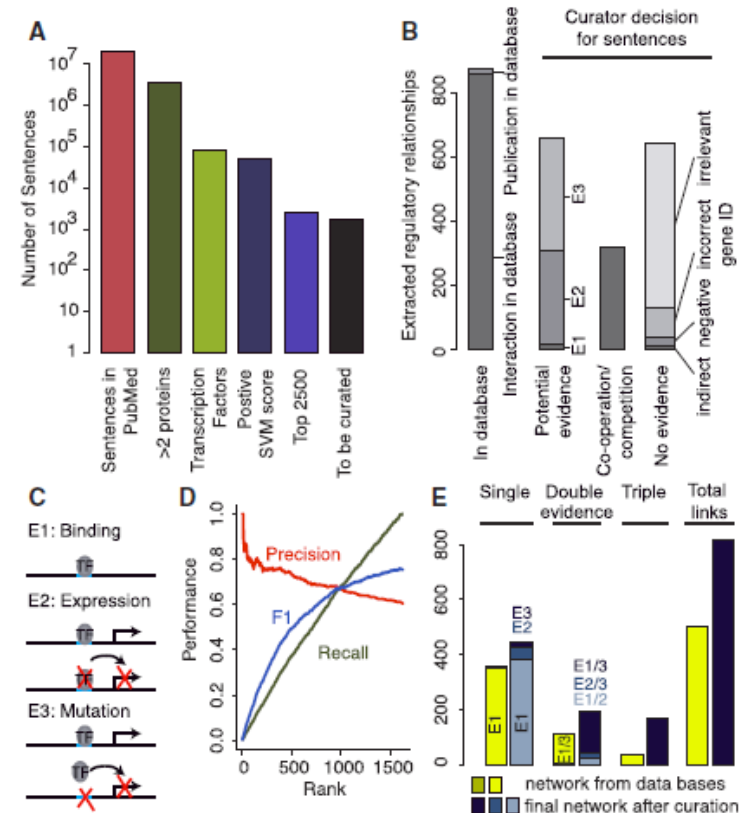


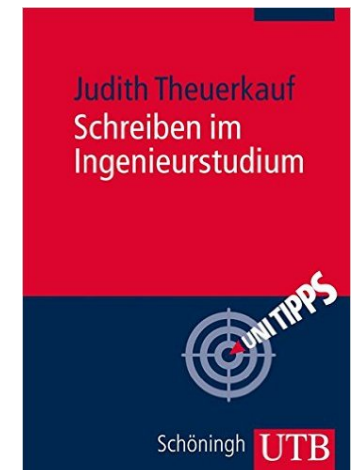
Fig. 3. Curation of regulatory interactions (A) Number of sentences that were considered in each of the steps of the pipeline. (B) Curator decisions for the 1625 sentences with highest rank. (C) The three evidence codes used for curation. (D) Precision, recall and F1-measure for manually curated sentences ranked by their confidence score. (E) Frequency of the different evidence levels for the existing relations in databases (yellow), or after full-text curation (blue)

Inhalt

- Inhalt
- Schreibstil
- Formatierung
- Bibliographie
- Zitieren / Plagiate
- Weitere Elemente
- Letzte Worte

Letzte Worte

- Lassen Sie Ihre Arbeit von mindestens einem **anderen Kommilitonen lesen** – Korrektheit und Verständlichkeit
- Prüfen Sie alle Punkte dieses Foliensatzes
 - Absätze sinnvoll gesetzt?
 - Literaturangaben vollständig?
 - Textuelle Aufzählungen durch Bullets ersetzt?
 - Einheitliches (langweiliges) Textbild?
 - Füllwörter alle entfernt?
 - ...
- Eine gute Arbeit wird **mehrmals umgeschrieben** – nehmen Sie sich die Zeit



Deliverable 3: First Written Text (4 Pages)

- Die 4-Seiten Fassung soll die Seminararbeit in Kurzform sein. Sie soll sowohl sehr sorgfältig formuliert als auch korrekt und umfassend im Sinne des Themas sein - aber es fehlen Details, für Beispiele wird kaum Platz sein, etc. Die Gliederung sollte aber im Grunde die gleiche wie bei der endgültigen Arbeit sein. Sinn dieser Arbeit ist es, dass Sie auf 4 Seiten gutes Schreiben und die Erfassung eines Themas üben können, ohne vor zu viel Text zurückzuschrecken. Idealerweise ist die 10-Seiten Fassung am Ende einfach die 4-Seitenfassung, ergänzt um Beweise, Messungen, Beispiele und ausführlichere Erklärungen. Das Feedback zur 4-Seiten Fassung erfolgt im individuellen Gespräch
- Die 4 Seiten sind inkl. Bibliographie, aber zusätzlich zu Inhaltsverzeichnis und Titelseite