Schreiben einer wissenschaftlichen Arbeit

Marina Sedinkina
Folien von Benjamin Roth
marina.sedinkina@campus.lmu.de
www.cis.lmu.de~/sedinkina

CIS LMU München

Allgemeines

- Viele allgemeine Hinweise zur Anfertigung von Bachelorarbeiten in anderen Fächern lassen sich nicht direkt auf die Computerlinguistik übertragen.
- Computerlinguistik: Bearbeitung einer Fragestellung enthält meist praktische (programmier-)Arbeit und deren Evaluation.
- Im Folgenden Vorschläge und Gedankenanstöße zu
 - Gliederung
 - Literaturrecherche
 - Schreibprozess
 - Sprachstil
 - Bewertungskriterien an die Arbeit
- Bitte auch mit dem jeweiligen Betreuer klären, was erwartet wird!

Gliederung

- Gliederung der Arbeit
- 2 Literaturrecherche
- Schreibstil
- 4 Schreibprozess
- Bewertungskriterien

• Hilft dem Schreibenden den Inhalt klar darzulegen

- Hilft dem Schreibenden den Inhalt klar darzulegen
- Logischer Aufbau:

- Hilft dem Schreibenden den Inhalt klar darzulegen
- Logischer Aufbau:
 - Motivierung vor Ausführung

- Hilft dem Schreibenden den Inhalt klar darzulegen
- Logischer Aufbau:
 - Motivierung vor Ausführung
 - Allgemeines vor Konkretem

- Hilft dem Schreibenden den Inhalt klar darzulegen
- Logischer Aufbau:
 - Motivierung vor Ausführung
 - Allgemeines vor Konkretem
 - Keine Wiederholung oder Redundanz (Ausnahme: Einleitungs- und Zusammenfassungsteile)

- Hilft dem Schreibenden den Inhalt klar darzulegen
- Logischer Aufbau:
 - Motivierung vor Ausführung
 - Allgemeines vor Konkretem
 - Keine Wiederholung oder Redundanz (Ausnahme: Einleitungs- und Zusammenfassungsteile)
- Hilft dem Leser sich schnell einen Überblick zu verschaffen

- Hilft dem Schreibenden den Inhalt klar darzulegen
- Logischer Aufbau:
 - Motivierung vor Ausführung
 - Allgemeines vor Konkretem
 - Keine Wiederholung oder Redundanz (Ausnahme: Einleitungs- und Zusammenfassungsteile)
- Hilft dem Leser sich schnell einen Überblick zu verschaffen
 - Worum geht es in der Arbeit?

- Hilft dem Schreibenden den Inhalt klar darzulegen
- Logischer Aufbau:
 - Motivierung vor Ausführung
 - Allgemeines vor Konkretem
 - Keine Wiederholung oder Redundanz (Ausnahme: Einleitungs- und Zusammenfassungsteile)
- Hilft dem Leser sich schnell einen Überblick zu verschaffen
 - Worum geht es in der Arbeit?
 - ▶ An welcher Stelle finde ich eine bestimmte Information?

- Hilft dem Schreibenden den Inhalt klar darzulegen
- Logischer Aufbau:
 - Motivierung vor Ausführung
 - Allgemeines vor Konkretem
 - Keine Wiederholung oder Redundanz (Ausnahme: Einleitungs- und Zusammenfassungsteile)
- Hilft dem Leser sich schnell einen Überblick zu verschaffen
 - Worum geht es in der Arbeit?
 - ▶ An welcher Stelle finde ich eine bestimmte Information?
- Gliederungspunkte und Unterpunkte

- Hilft dem Schreibenden den Inhalt klar darzulegen
- Logischer Aufbau:
 - Motivierung vor Ausführung
 - Allgemeines vor Konkretem
 - Keine Wiederholung oder Redundanz (Ausnahme: Einleitungs- und Zusammenfassungsteile)
- Hilft dem Leser sich schnell einen Überblick zu verschaffen
 - Worum geht es in der Arbeit?
 - ▶ An welcher Stelle finde ich eine bestimmte Information?
- Gliederungspunkte und Unterpunkte
 - ▶ Überschrift: Kurzcharakterisierung des Inhalts im Nominalstil

- Hilft dem Schreibenden den Inhalt klar darzulegen
- Logischer Aufbau:
 - Motivierung vor Ausführung
 - Allgemeines vor Konkretem
 - Keine Wiederholung oder Redundanz (Ausnahme: Einleitungs- und Zusammenfassungsteile)
- Hilft dem Leser sich schnell einen Überblick zu verschaffen
 - Worum geht es in der Arbeit?
 - ▶ An welcher Stelle finde ich eine bestimmte Information?
- Gliederungspunkte und Unterpunkte
 - ▶ Überschrift: Kurzcharakterisierung des Inhalts im Nominalstil
 - ▶ Richtwert: Pro kleinster Gliederungseinheit $\frac{1}{2}$ bis 2 Seiten Text.

(Im konkreten Fall werden natürlich andere Kapitelüberschriften gewählt!)

Murzübersicht über die Arbeit (Einleitung)

- Kurzübersicht über die Arbeit (Einleitung)
- 2 Thematischer Hintergrund und Definition der Fragestellung

- Kurzübersicht über die Arbeit (Einleitung)
- Thematischer Hintergrund und Definition der Fragestellung
 - Einführung in das Themenfeld (beinhaltet Darlegung seiner Relevanz)

- Kurzübersicht über die Arbeit (Einleitung)
- Thematischer Hintergrund und Definition der Fragestellung
 - Einführung in das Themenfeld (beinhaltet Darlegung seiner Relevanz)
 - Formulieren der Forschungsfrage / These (besonderes Augenmerk auf Relevanz der Fragestellung!)

- Kurzübersicht über die Arbeit (Einleitung)
- Thematischer Hintergrund und Definition der Fragestellung
 - Einführung in das Themenfeld (beinhaltet Darlegung seiner Relevanz)
 - Formulieren der Forschungsfrage / These (besonderes Augenmerk auf Relevanz der Fragestellung!)
- Übersicht über relevante Forschungliteratur zur Fragestellung

- Kurzübersicht über die Arbeit (Einleitung)
- Thematischer Hintergrund und Definition der Fragestellung
 - Einführung in das Themenfeld (beinhaltet Darlegung seiner Relevanz)
 - Formulieren der Forschungsfrage / These (besonderes Augenmerk auf Relevanz der Fragestellung!)
- Übersicht über relevante Forschungliteratur zur Fragestellung
- Beschreibung des eigenen Forschungsbeitrags

- Kurzübersicht über die Arbeit (Einleitung)
- Thematischer Hintergrund und Definition der Fragestellung
 - Einführung in das Themenfeld (beinhaltet Darlegung seiner Relevanz)
 - Formulieren der Forschungsfrage / These (besonderes Augenmerk auf Relevanz der Fragestellung!)
- Übersicht über relevante Forschungliteratur zur Fragestellung
- Beschreibung des eigenen Forschungsbeitrags
 - Motivation des Ansatzes / Theoretische Uberlegungen

- Kurzübersicht über die Arbeit (Einleitung)
- Thematischer Hintergrund und Definition der Fragestellung
 - Einführung in das Themenfeld (beinhaltet Darlegung seiner Relevanz)
 - Formulieren der Forschungsfrage / These (besonderes Augenmerk auf Relevanz der Fragestellung!)
- Übersicht über relevante Forschungliteratur zur Fragestellung
- Beschreibung des eigenen Forschungsbeitrags
 - Motivation des Ansatzes / Theoretische Überlegungen
 - Beschreibung des selbst entwickelten Verfahrens / Algorithmus' / Systems / der Versuchsanordnung / der Ressource

- Kurzübersicht über die Arbeit (Einleitung)
- Thematischer Hintergrund und Definition der Fragestellung
 - Einführung in das Themenfeld (beinhaltet Darlegung seiner Relevanz)
 - Formulieren der Forschungsfrage / These (besonderes Augenmerk auf Relevanz der Fragestellung!)
- Übersicht über relevante Forschungliteratur zur Fragestellung
- Beschreibung des eigenen Forschungsbeitrags
 - Motivation des Ansatzes / Theoretische Überlegungen
 - Beschreibung des selbst entwickelten Verfahrens / Algorithmus' / Systems / der Versuchsanordnung / der Ressource
 - Quantitativer Teil / Experimente

- Kurzübersicht über die Arbeit (Einleitung)
- Thematischer Hintergrund und Definition der Fragestellung
 - Einführung in das Themenfeld (beinhaltet Darlegung seiner Relevanz)
 - Formulieren der Forschungsfrage / These (besonderes Augenmerk auf Relevanz der Fragestellung!)
- Übersicht über relevante Forschungliteratur zur Fragestellung
- Beschreibung des eigenen Forschungsbeitrags
 - Motivation des Ansatzes / Theoretische Überlegungen
 - Beschreibung des selbst entwickelten Verfahrens / Algorithmus' / Systems / der Versuchsanordnung / der Ressource
 - Quantitativer Teil / Experimente
 - Beschreibung der Daten

- Kurzübersicht über die Arbeit (Einleitung)
- Thematischer Hintergrund und Definition der Fragestellung
 - Einführung in das Themenfeld (beinhaltet Darlegung seiner Relevanz)
 - Formulieren der Forschungsfrage / These (besonderes Augenmerk auf Relevanz der Fragestellung!)
- Übersicht über relevante Forschungliteratur zur Fragestellung
- Beschreibung des eigenen Forschungsbeitrags
 - Motivation des Ansatzes / Theoretische Überlegungen
 - Beschreibung des selbst entwickelten Verfahrens / Algorithmus' / Systems / der Versuchsanordnung / der Ressource
 - Quantitativer Teil / Experimente
 - Beschreibung der Daten
 - Details zur Versuchsdurchführung

- Kurzübersicht über die Arbeit (Einleitung)
- Thematischer Hintergrund und Definition der Fragestellung
 - Einführung in das Themenfeld (beinhaltet Darlegung seiner Relevanz)
 - Formulieren der Forschungsfrage / These (besonderes Augenmerk auf Relevanz der Fragestellung!)
- Übersicht über relevante Forschungliteratur zur Fragestellung
- Beschreibung des eigenen Forschungsbeitrags
 - Motivation des Ansatzes / Theoretische Überlegungen
 - Beschreibung des selbst entwickelten Verfahrens / Algorithmus' / Systems / der Versuchsanordnung / der Ressource
 - Quantitativer Teil / Experimente
 - Beschreibung der Daten
 - Details zur Versuchsdurchführung
 - Versuchsergebnisse / Signifikanztests

- Kurzübersicht über die Arbeit (Einleitung)
- Thematischer Hintergrund und Definition der Fragestellung
 - Einführung in das Themenfeld (beinhaltet Darlegung seiner Relevanz)
 - Formulieren der Forschungsfrage / These (besonderes Augenmerk auf Relevanz der Fragestellung!)
- Übersicht über relevante Forschungliteratur zur Fragestellung
- Beschreibung des eigenen Forschungsbeitrags
 - Motivation des Ansatzes / Theoretische Überlegungen
 - Beschreibung des selbst entwickelten Verfahrens / Algorithmus' / Systems / der Versuchsanordnung / der Ressource
 - Quantitativer Teil / Experimente
 - Beschreibung der Daten
 - Details zur Versuchsdurchführung
 - Versuchsergebnisse / Signifikanztests
 - Diskussion der Ergebnisse (qualitativ)

(Im konkreten Fall werden natürlich andere Kapitelüberschriften gewählt!)

- Murzübersicht über die Arbeit (Einleitung)
- Thematischer Hintergrund und Definition der Fragestellung
 - Einführung in das Themenfeld (beinhaltet Darlegung seiner Relevanz)
 - Formulieren der Forschungsfrage / These (besonderes Augenmerk auf Relevanz der Fragestellung!)
- Übersicht über relevante Forschungliteratur zur Fragestellung
- Beschreibung des eigenen Forschungsbeitrags
 - Motivation des Ansatzes / Theoretische Überlegungen
 - Beschreibung des selbst entwickelten Verfahrens / Algorithmus' / Systems / der Versuchsanordnung / der Ressource
 - Quantitativer Teil / Experimente
 - Beschreibung der Daten
 - Details zur Versuchsdurchführung
 - Versuchsergebnisse / Signifikanztests
 - Diskussion der Ergebnisse (qualitativ)
- Zusammenfassung

5 / 30

Kapitel: Einleitung / Kurzübersicht

- Kurz! $(\frac{1}{2} 1 \text{ Seite})$
- Gibt dem Leser einen Überblick über den Rest der Arbeit
- Alternativer Gliederungsaufbau: Kurzübersicht, Thematischer Hintergrund und Fragestellung bilden zusammen das Einleitungskapitel.
 - (D.h. die ersten zwei Punkte werden zusammengelegt)

Kapitel: Thematischer Hintergrund und Definition der Fragestellung

- Einführung in das Themenfeld
 - Imaginierter Leser: Computerlinguist (Bachelor), dessen Spezialgebiet ein anderes Thema ist.
 - ► Herausstellen der Relevanz!

Kapitel: Thematischer Hintergrund und Definition der Fragestellung

- Formulieren der Forschungsfrage/These
 - ► Es ist extrem schwierig (unrealistisch) in einer Bachelorarbeit einen relevant Forschungsbeitrag zu leisten...
 - ▶ Dennoch sollte es versucht werden!
 - ► Herausstellen der Relevanz: Lücken im Stand der Forschung? Lösung eines praktischen Problems? ...
 - Selbst sehr praktische Themen sollten als Fragestellung formuliert werden:
 - "... soll untersucht werden, inwieweit reguläre Ausdrücke ein geeignetes Verfahren zum Erkennen von Datumsangaben darstellen ..."

Kapitel: Relevante Literatur

- Relevante Literatur sollte kurz zusammengefasst und in den Gesamtkontext eingeordnet werden.
- Sowohl "Klassiker" als auch "Cutting Edge" können relevant sein.
- Später noch Tipps zu folgenden Fragen:
 - Wie finde ich Literatur?
 - Wie beurteile ich die Qualität?
 - ► Wie lese ich / verschaffe ich mir einen Uberblick über gefundene Literatur?

Kapitel: Eigener Forschungsbeitrag

- Richtwert: $\frac{2}{3}$ der Arbeit
- Beschreibung des eigenen Beitrags, z.B.
 - Software/System
 - Resource
 - Formalismus
 - Beweis
 - ► Konzept/Theorie
 - **.**..
- Der eigene Beitrag sollte quantitativ untersucht werden
 - Statistiken (Data Overview)
 - Experimente
 - Zahlen zu jedem Experiment (f-score, accuracy etc.)
- Die Ergebnisse führen zu einer qualitativen Charakterisierung
 - Einschränkungen und Potentiale
 - Beispiele aus den Daten
- Bei bestimmten Forschungsthemen (mathematischer Beweis ...) wird der eigene Beitrag entsprechend anders untersucht.

Kapitel: Zusammenfassung

- Ähnlich der Einleitung, allerdings kann Lektüre der Arbeit bereits vorausgesetzt werden.
- Betonung der wichtigsten Ergebnisse.
- Einordnung in Gesamtkontext und evtl. Ausblick.
- Kurz: ~ 1 Seite

Gliederung

- Gliederung der Arbeit
- 2 Literaturrecherche
- Schreibstil
- 4 Schreibprozess
- Bewertungskriterien

Literatur finden: Einstiegshilfen

- Literaturhinweise vom Betreuer der Arbeit.
- Literaturhinweise von anderen Experten.
- Lehrbücher (in Bibliothek).
- Wichtigste Konferenzen und Journale.
- Internetsuche
 - Seiten von Wissenschaftlern
 - Tutorials
 - ▶ Blogs, Twitter
 - **...**
- Google Scholar
- Aus Google Scholar direkt zum Volltext: http://www.ub.uni-muenchen.de/suchen/google-scholar/index.html

Literatur finden: Einstiegshilfen

- Empfehlung: Einstiegsliteratur hauptsächlich finden durch
 - Literaturhinweise vom Betreuer der Arbeit
 - ► Top-Konferenzen der letzten Jahre
- Andere Wege je nach Präferenz zusätzlich nutzen.

Wichtigste Konferenzen und Journale

• Konferenzen:

- ACL (Association for Computational Linguistics) http://aclweb.org/anthology/
- ► EMNLP (Empirical Methods for Natural Language Unterstanding
- NAACL, EACL, IJCNLP, COLING, CoNLL

Journale

- ► TACL (Transaction of the Association for Computational Linguistics) https://www.transacl.org/ojs/index.php/tacl/index
- ► Computational Linguistics (MIT Press)
 https://www.mitpressjournals.org/toc/coli/44/1

Literatur Schnellüberblick und Auswahl

- Man kann sich in sehr kurzer Zeit einen Überblick über ein große Anzahl von Papers verschaffen.
- Ein paar wenige Papers mit der höchsten Relevanz können später genau gelesen werden.
- Querlesen: ∼ 3-5 Minuten pro Paper
- Wesentliche Information enthalten in:
 - Abstract
 - Aufbau
 - Abbildungen
 - Summary

Literatur Schnellüberblick und Auswahl: Querlesen

- Querlesen: ∼ 3-5 Minuten pro Paper
- Man legt eine Tabelle mit einer Zeile pro quergelesenem Paper an.
- Spalten der Tabelle:
 - Titel, Autor, Jahr
 - Relevanz (erster Eindruck), Werte von 1-5
 - Zusammenfassung (ein Satz)
 - Optional: Neue Ideen (ein Satz)
- Mehr Details:
 - J.S. Yi. "QnDReview: Read 100 CHI Papers in 7 Hours." CHI 2014

Literaturliste erweitern

- Nach dem Identifizieren einer kleinen Menge relevanter Literatur kann davon ausgehend weiter gesucht werden
 - Darin zitierte Literatur
 - Zitierende Literatur (Google Scholar, MS Academic (https://academic.microsoft.com/), ...)
 - Webseiten von Autoren
 - ★ Weitere Papers der Autoren
 - Spezialkonferenzen / Workshops

Qualität der Literatur einschätzen

- Am Ende zählt der Inhalt.
- Bei einem neuen Thema kann man diesen jedoch noch nicht perfekt beurteilen.
- Heuristiken:
 - Welche Konferenz/Journal? (Rankings, Impact factors)
 - Welche Universität?
 - Wieviele Zitationen?
 - Welche (Co-)Autoren?

Gliederung

- Gliederung der Arbeit
- 2 Literaturrecherche
- Schreibstil
- 4 Schreibprozess
- Bewertungskriterien

Schreibstil

- Sachlich "Leider verhält es sich so, dass ...",
 "Diese These stützt glücklicherweise meine Ausführung ..."
- Eindeutige Definition und konsistente Verwendung der Terminologie: enweder immer *SVM* oder immer *SVMs*
- Klare Sätze (im Zweifel zwei kurze statt ein langer Satz).
- Erste Person (ich, wir) vermeiden.
- Tempus: Hauptsächlich Präsens, in bestimmten Kontexten Präteritum

Schreibstil: Tempus

- Präsens, fast durchgehend
 - ► Forschungstand, Allgemeingültige Tatsachen
 - Forschungsfrage
 - ▶ Diskussion der Ergebnisse, Fazit
- Präteritum, in Ausnahmefällen
 - Bei Literaturhinweisen, wenn diese als vergangenes Ereignis dargestellt werden:
 - "Hearst (1992) entwickelte das erste Verfahren zur Extraktion von taxonomische Relationen aus Texten."
 - Aber: "Bei Verfahren, die nicht auf maschinellem Lernen beruhen (Hearst, 1992), werden keine Trainingsdaten benötigt."
 - Bei Beschreibung der tatsächlichen Experimentendurchführung
 "Es wurden drei Modelle mit Hyperparametern 1, 10 und 100 trainiert."

 Aber: "Tabelle 1 zeigt die Ergebnisse für die Hyperparameter 1, 10 und 100"

Gliederung

- Gliederung der Arbeit
- 2 Literaturrecherche
- Schreibstil
- 4 Schreibprozess
- Bewertungskriterien

Schreibprozess

- "Traditionelles" Vorgehen
 - Erst Gliederung anfertigen
 - Dann Einzelpunkte mit Inhalt füllen
- Kritik an traditionellem Vorgehen
 - Kann Schreibfluss behindern
 - Ursprüngliche Gliederung entspricht ohnehin nicht dem Endergebnis
- "Alternatives" Vorgehen (nach Bolker)
 - ► Erst Textmaterial sammeln
 - Dann Gliedern
- J. Bolker: Writing Your Dissertation in Fifteen Minutes a Day
- Argumente f
 ür alternative Vorgehensweise
 - Ideen zur Gliederung kommen wenn Material vorliegt.
 - Ideen zu Formulierungen kommen beim Schreiben.

Schreibprozess nach Bolker

- Wichtigstes Prinzip ist, in den Schreibfluss zu kommen.
- Technik: über irgendetwas schreiben, selbst wenn es nichts mit dem Thema zu tun hat.
 - ▶ Irrelevanter oder Redundanter Text kann später wieder entfernt werden.
 - ▶ Die Kommentarfunktion (z.B. von Latex) kann dies unterstützen.
- Wenn genug Material da ist, dieses in mehreren Zyklen
 - ▶ immer weiter verfeinern und überarbeiten
- Am Ende
 - Kollegen und Freunden bitten Korrektur zu lesen
 - Rechtschreibprogramm nicht vergessen!

Schreibprozess

- Natürlich sind viele Kombinationen aus <u>"traditionellem"</u> und "alternativem" Ansatz denkbar.
- Zum Beispiel:
 - Zunächst wird eine Arbeitsgliederung erstellt.
 - ▶ Diese wird ständig angepasst, oder bei Bedarf ignoriert.
 - ▶ Die Teile werden nicht in Reihenfolge der Gliederung abgearbeitet. Tipp: zuerst mit "related work" anfangen
 - ▶ Am Ende gründliche Überarbeitung von Text und Gliederung.

Gliederung

- Gliederung der Arbeit
- 2 Literaturrecherche
- Schreibstil
- 4 Schreibprozess
- Bewertungskriterien

Bewertungskriterien

(Aufzählung ist nicht vollständig - mit jeweiligem Betreuer sprechen!)

- Sind Forschungsfrage und zu untersuchende Thesen klar formuliert?
- Aufbau und innere Struktur
 - ▶ Ist die Arbeit logisch aufgebaut?
 - ▶ Ist die Gliederung in Bezug auf das Thema aussagefähig?
- Aufarbeitung des Forschungsstands
 - Wird die wichtigste Literatur zum Thema vorgestellt?
 - Wird sie kritisch gewürdigt und richtig eingeordnet?
 - Ist die Auswahl relevant?
 - Eigenständigkeit bei der Literatursuche.
- Methoden und Analysen
 - Ist die gewählte Vorgehensweise adäquat?
 - ► Können Aussagen quantitativ belegt werden?
 - Eigenständigkeit bei Auswahl des Verfahrens (oder Modifikation eines Verfahrens)

Bewertungskriterien

- Argumentation und Interpretation
 - Werden eigenständige Beobachtungen gemacht und korrekte Schlussfolgerungen gezogen?
 - Wird der Gültigkeitsbereich von Aussagen klar definiert und abgegrenzt?
- Verständlichkeit, Sprache und Stil
 - Ist die Arbeit in klarer und sachlicher Sprache verfasst?
 - Wird Verständlichkeit durch Beispiele und Abbildungen angestrebt?
 - Ist Terminologie korrekt eingeführt und konsistent verwendet?
 - ► Fehlerfrei in Bezug auf Rechtschreibung, Grammatik, Zeichensetzung?
- Sind die formalen Kriterien eingehalten? (Deckblatt, eidesstattliche Erklärung, ...)

• Noch Fragen?