

**Automatische Evaluation  
semantischer Kohärenz  
bei Topic Models**

Projektarbeit: Freies Projekt im digitalen Bereich  
Aufbaumodul Projekt  
Sommersemester 2017  
Julius-Maximilians-Universität Würzburg  
Philosophische Fakultät  
Institut für Deutsche Philologie  
Lehrstuhl für Computerphilologie und  
Neuere Deutsche Literaturgeschichte

Betreuer: Prof. Dr. Fotis Jannidis

Severin Simmler  
Sophienstraße 6  
97072 Würzburg  
severin.simmler@stud-mail.uni-wuerzburg.de  
Matrikel Nr.: 2028090  
Studienfach: Digital Humanities (B.A.)  
Fachsemester: 6

Würzburg, 16.05.2017

# INHALT

1. Einführung .....	3
2. Theorie: Grundbegriffe und Definitionen .....	3
2.1 Latent Dirichlet allocation .....	3
2.2 Kohärenzmaße .....	4
2.2.1 PMI (UCI) .....	4
2.2.1 PMI (UMass) .....	5
3. Verwendete Korpora .....	5
4. Praxis: Konzept und Implementierung .....	5
4.1 Topics – Easy Topic Modeling .....	6
4.2 Datenmodellierung .....	6
4.3 class Preprocessing .....	7
4.3 class Measures .....	8
4.4 class Evaluation .....	8
4.5 Experimente .....	9
4.5.1 Deutschsprachige Erzählungen .....	9
4.5.2 Röder Topics .....	10
4.5.3 Wikipedia .....	11
5. Fazit und Ausblick .....	11
6. Literaturverzeichnis .....	12
7. Selbstständigkeitserklärung .....	13

## ABBILDUNGS-, CODE- UND TABELLENVERZEICHNIS

Abbildung 1: UML-Diagramm des Moduls .....	7
Pseudocode 1: segment_topics .....	7
Pseudocode 2: calculate_occurences .....	8
Pseudocode 3: pmi_umass .....	8
Pseudocode 4: pmi_uci .....	8
Pseudocode 5: calculate_umass or calculate_uci .....	8
Tabelle 1: Kracht Topics (10 × 10) und Kracht Korpus .....	9
Tabelle 3: Kracht Topics (10 × 10) und deutsches Wikipedia Korpus .....	10
Tabelle 2: Röder Topics (5 × 100) und englisches Wikipedia Korpus .....	10
Tabelle 4: Wikipedia Topics (10 × 10) und deutsches Wikipedia Korpus .....	11

## 1. Einführung

Kohärenz hilft uns zu beschreiben, wie nützlich Informationen empfunden werden. Zweifellos ist  $S = \{\text{all ravens are black, this bird is a raven, this bird is black}\}$  kohärenter als  $S' = \{\text{this chair is brown, electrons are negatively charged, today is Thursday}\}$  (BonJour 1985: 96). Alle drei Komponenten von  $S$  unterstützen sich semantisch gegenseitig, wohingegen  $S'$  aus drei (semantisch) unabhängigen Teilen besteht. Deswegen wird  $S$  nützlicher als  $S'$  empfunden.

Topic Models extrahieren latente semantische Strukturen aus unstrukturierter Daten. Diese Strukturen werden durch Topics beschrieben, die aus  $n$  Worten bestehen. Es gibt keine Garantie, dass ein Model nützliche – oder genauer gesagt kohärente – Topics generiert. In der vorliegenden Arbeit will ich zwei Kohärenzmaße vorstellen, mit denen Topic Models quantitativ evaluiert werden können. Hierbei gilt analog zum oben genannten Beispiel: je kohärenter ein Topic, desto nützlicher ist es.

Zunächst will ich in aller Kürze auf die theoretischen Grundlagen des Topic Modeling eingehen (2.1), um anschließend die Kohärenzmaße vorzustellen (2.2). Nach einer Beschreibung der verwendeten Datensätze (3.), wird im praktischen Teil (4.) das Konzept und die Implementierung eines Python Moduls besprochen. Nach der Diskussion aller Experimente und Ergebnisse (4.5), schließe ich in einem kurzen Fazit mit Ausblick auf zukünftige Pläne (6.) ab.

## 2. Theorie: Grundbegriffe und Definitionen

Im folgenden Teil sollen theoretische Grundlagen knapp definiert und nur kurz ausgeführt werden, da der Schwerpunkt dieser Arbeit auf den praktischen Teil gelegt ist.

### 2.1 Latent Dirichlet allocation

Blei et al. beschreiben *latent Dirichlet allocation* (LDA) als ein generatives Wahrscheinlichkeitsmodell, das jedes Dokument eines Korpus als „finite mixture over an underlying set of topics“ (2003: 993) modelliert. Das Korpus besteht hierbei aus einer Sammlung von  $M$  Dokumenten  $D = \{d_1, d_2, \dots, d_M\}$  und jedes Dokument aus  $N$  Worten  $d = \{w_1, w_2, \dots, w_N\}$ . Jedes einzigartige Wort  $V$  in  $D$  wird durch einen

fortlaufenden Index  $\{1, \dots, V\}$  dargestellt (Blei et al. 2003: 995) und entspricht somit einem *bag-of-words* (Harris 1954).

In den folgenden Experimenten wird ein Topic  $t = \{w_1, w_2, \dots, w_{10}\}$  mit den wahrscheinlichsten Tokens für  $t$  (,Keys') absteigend sortiert beschrieben. Dabei wurde für jedes Dokument in  $n$  Iterationen (i) zufällig eine Verteilung über die Topics generiert, (ii) zufällig ein Topic ausgewählt und (iii) diesem zufällig ein Wort aus  $D$  zugeteilt (Blei 2012: 78).

## 2.2 Kohärenzmaße

Je mehr sich einzelne Komponenten einer Wortgruppe gegenseitig unterstützen, desto einfacher ist es für Menschen einen Sinnzusammenhang herzustellen. Je besser eine Wortgruppe zu interpretieren ist, desto höher ist ihr Grad an Kohärenz. Kohärenzmaße sollen im Kontext von Topic Modeling ein Werkzeug sein, um die semantische Qualität von Topics automatisch zu bewerten.

$C_{UMass}$  (2.2.1) soll das Korpus einbeziehen, mit dem das Topic Model trainiert wurde,  $C_{UCI}$  (2.2.2) evaluiert die Topics mithilfe eines externen, unbekanntes Korpus. In den folgenden Experimenten werden allerdings beide Maße sowohl intrinsisch als auch extrinsisch angewandt.

### 2.2.1 PMI (UCI)

Newman et al. (2010) führen ein Kohärenzmaß basierend auf PMI (Church et al. 1990) ein. Dabei werden die Wahrscheinlichkeiten verglichen, dass  $w_i$  und  $w_j$  jeweils zusammen (Zähler) und unabhängig voneinander (Nenner) auftreten und ist definiert als:

$$C_{UCI} = \log \frac{P(w_i, w_j) + \epsilon}{P(w_i) \cdot P(w_j)}$$

$\epsilon$  verhindert einen Logarithmus von 0 und liefert bei einem Wert  $< 1$  bessere Werte (Stevens et al. 2012: 955, vgl. auch Röder et al. 2015: 3).

Aletras und Stevenson (2013) konnten mit der normalisierten Variante bessere Ergebnisse erzielen:

$$C_{UCI}^{normalized} = \frac{\log \frac{P(w_i, w_j) + \epsilon}{P(w_i) \cdot P(w_j)}}{-\log (P(w_i, w_j))}$$

### 2.2.1 PMI (UMass)

Mimno et al. (2011) definieren ein Kohärenzmaß als:

$$C_{UMass} = \log \frac{P(w_i, w_j) + \epsilon}{P(w_j)}$$

$C_{UMass}$  basiert auf PMI, Mimno et al. konnten allerdings mit ihrer Variante bessere Ergebnisse erzielen, da es mehr auf „the conditional probability of each word given the each of the higher-ranked words in the topic“ (2011: 266) als auf die ursprüngliche Intention von PMI, die gemeinsame Wahrscheinlichkeit durch das Produkt der einzelnen Wahrscheinlichkeiten zu teilen.

## 3. Verwendete Korpora

Zur Verfügung standen (i) eine Sammlung von zwanzig deutschsprachigen Erzählungen (Kracht 2002), die 9.179 einzigartige Tokens enthalten, (ii) eine Sammlung von 1.023.005 Artikeln (Dokumentlänge > 30 Tokens) aus der deutschen Wikipedia<sup>1</sup>, die 6.256.302 einzigartige Tokens enthalten, (iii) eine Sammlung von 125.366 Artikeln (Sparte: Film) aus der englischen Wikipedia<sup>2</sup>, die 827.794 einzigartige Tokens beinhalten und (iv) 255 zufällig ausgewählte Artikel aus der deutschen Wikipedia<sup>1</sup> mit 71.112 einzigartigen Tokens.

Alle Texte lagen als *plain text* vor, wurden tokenisiert<sup>3</sup>, bereinigt<sup>4</sup> und in das Textdatenmodell von DARIAH überführt (siehe 4.1 Datenmodellierung).

## 4. Praxis: Konzept und Implementierung

Im folgenden Teil werden die bisher verfügbaren Module kurz erläutert, um anschließend über die Datenmodellierung zur Beschreibung des im Rahmen dieser Arbeit entwickelten Moduls zu gelangen.

---

<sup>1</sup> Basierend auf einem XML-Dump (<https://dumps.wikimedia.org/dewiki/latest>) vom 21. April 2017 und mithilfe von WikiExtractor (<https://github.com/attardi/wikiextractor>) als *plain text* extrahiert.

<sup>2</sup> Freundlicherweise zur Verfügung gestellt von Michael Röder, Universität Leipzig (<http://139.18.2.164/mroeder/palmetto/datasets/>)

<sup>3</sup> Mit einem regulären Ausdruck ( $\backslash p\{L\}+\backslash p\{P\}?\backslash p\{L\}+$ )

<sup>4</sup> Stoppwörter wurden entfernt und alle Tokens kleingeschrieben.

## 4.1 Topics – Easy Topic Modeling

Im Rahmen von DARIAH-DE (Cluster 5: Quantitative Datenanalyse) entsteht eine Python Bibliothek<sup>5</sup>, die Module zur Verarbeitung großer und unstrukturierter Textkorpora zur Verfügung stellen soll. Bisher existiert (i) ein Modul zur Datenvorbereitung, (ii) ein Modul, das zwei populäre LDA Implementierungen<sup>6</sup> aufruft und (iii) ein Modul zur Visualisierung des Topic Models. Das hier vorgestellte Modul soll in die Bibliothek integriert werden.

## 4.2 Datenmodellierung

Ein Korpus für DARIAH Easy Topic Modeling wird im Zuge der Vorverarbeitung in ein *bag-of-words* überführt; jedem Dokument und jedem einzigartigen Wort wird dabei eine Zahl zugewiesen und das Korpus in einer Matrix durch die jeweils zugewiesenen Zahlen beschrieben.

Das Beispielkorpus [["Ich mag das Goethe-Institut"]] wird also mit {'ich': 1, 'mag': 2, 'das': 3, 'goethe-institut': 4} und {'dokument.txt': 1} folgendermaßen repräsentiert:

doc_id	type_id	freq
1	1	1
	2	1
	3	1
	4	1

Das Evaluationsmodul schließt an dieses Modell an und wurde objektorientiert in Python 3 implementiert. Einen ersten Überblick über Struktur und Umfang soll Abbildung 1 geben.

Die Klasse `Evaluation` ist eine Subklasse von `Preprocessing` und `Measures`. Sie wird instanziiert mit dem DARIAH *bag-of-words* (`DataFrame`), den Topics (`DataFrame`) und einem Dictionary der Form `{token: id}`.

---

<sup>5</sup> <<https://github.com/DARIAH-DE/Topics>>

<sup>6</sup> MALLET (McCallum 2002) und Gensim (Řehůřek 2009)

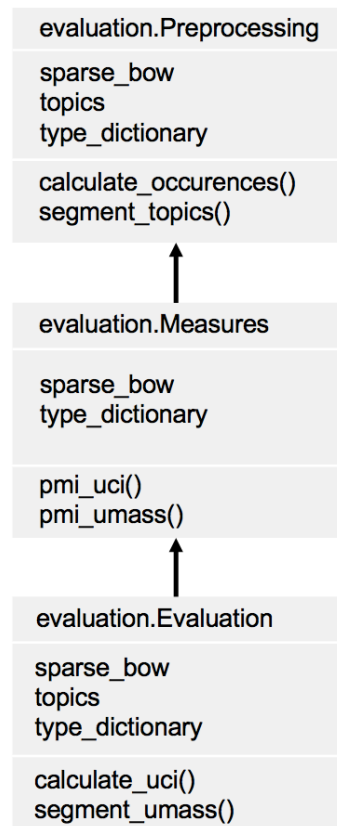


Abbildung 1: UML-Diagramm des Moduls

### 4.3 class Preprocessing

Die Klasse `Preprocessing` bereitet die Topics für die weitere Verarbeitung vor. Die Funktion `segment_topics` zerlegt die Topics und permutiert bzw. kombiniert die einzelnen Keys zu Bigrammen:

```

1  def segment_topics:
2      for topic in topics:
3          if permutation:
4              permutation(topic)
5          else:
6              combination(topic)
  
```

Pseudocode 1: `segment_topics`

Die folgende Funktion erwartet ein Set aller einzigartigen Topic Keys, für die alle Dokumente gezählt werden, in denen sie vorkommen. Basierend darauf können später auch sämtliche Kookkurrenzen ermittelt werden.

```

1 def calculate_occurences:
2     for key in set of keys:
3         for document in corpus:
4             if key in document:
5                 save or update {key, document id}

```

Pseudocode 2: calculate\_occurences

### 4.3 class Measures

Measures beinhaltet zwei Funktionen, die für ein Wortpaar sowohl PMI (UMass):

```

1 def pmi_umass:
2     numerator = freq bigram + e / corpus size
3     denominator = freq bigram[1] + e / corpus size
4     log(numerator / denominator)
5

```

Pseudocode 3: pmi\_umass

als auch PMI (UCI) berechnen:

```

1 def pmi_uci:
2     numerator = freq bigram + e / corpus size
3     denominator = (freq bigram[0] + e / corpus size) *
4                 (freq bigram[1] + e / corpus size)
5     if normalize:
6         log(numerator / denominator) / -log(numerator)
7     else:
8         log(numerator / denominator)

```

Pseudocode 4: pmi\_uci

### 4.4 class Evaluation

Schließlich bietet die Klasse Evaluation die Möglichkeit alle Topics eines Modells zu evaluieren:

```

1 def calculate_umass or calculate_uci:
2     segment topics
3     calculate occurences
4     for topic in segmented topics:
5         for bigram in topic:
6             save or update pmi(bigram)
7             normalize * mean(pmi scores)

```

Pseudocode 5: calculate\_umass or calculate\_uci

Mit  $\frac{2}{N \cdot (N-1)}$  und  $N = |t|$  wird der PMI pro Topic normalisiert (Z. 7).



## 4.5 Experimente

Die menschliche Bewertung von Topics wird als Goldstandard behandelt. Das Ziel ist demnach eine möglichst hohe Korrelation der Kohärenzwerte mit dem Goldstandard zu erreichen (vgl. Newman et al. 2010 oder Röder et al. 2015). Aus diesem Grund haben im Rahmen dieser Arbeit zwölf Personen über einen Onlinefragebogen<sup>7</sup> 20 Topics auf einer Skala von 0 (sinnlos) bis 2 (sinnvoll) bewertet (vgl. Newman et al. 2010). Für jedes bewertete Topic wurde der Mittelwert ermittelt.

### 4.5.1 Deutschsprachige Erzählungen

Aus zwanzig deutschsprachigen Erzählungen (Kracht 2002) wurden in 50.000 Iterationen 20 Topics generiert<sup>8</sup>. Tabelle 1 gibt einen Überblick über die Ergebnisse der ersten Experimente:

	<b>Zeit</b>	<b>Kohärenz (Median)</b>	<b>Spearman- Korrelation</b>	<b>p-Wert</b>
<b>UMass</b>	4 s	-0.049	0.385	0.094
<b>UMass (normalisiert)</b>	5 s	1.353	-0.273	0.245
<b>UCI</b>	1 s	-0.085	0.138	0.563
<b>UCI (normalisiert)</b>	1 s	3.691	-0.181	0.444

Tabelle 1: Kracht Topics (10 × 10) und Kracht Korpus

Alle p-Werte liegen deutlich über  $\alpha = 0.05$ . Es ist davon auszugehen, dass keine Zusammenhänge bestehen.

Tabelle 3 fasst die Ergebnisse der Experimente mit denselben Topics und 1.023.005 deutschen Wikipedia Artikeln zusammen:

	<b>Zeit</b>	<b>Kohärenz (Median)</b>	<b>Spearman- Korrelation</b>	<b>p-Wert</b>
<b>UMass</b>	25 min 47 s	-0.116	0.249	0.291

<sup>7</sup> Abrufbar über <https://goo.gl/forms/cd7TrRhpGiWYn6dm1>

<sup>8</sup> Topic Modeling mit MALLET (McCallum 2002).

<b>UMass (normalisiert)</b>	30 min 23 s	37957.154	-0.311	0.183
<b>UCI</b>	13 min 41 s	0.306	0.221	0.348
<b>UCI (norma- lisiert)</b>	13 min 22 s	82836.511	-0.157	0.508

Tabelle 3: Kracht Topics (10 × 10) und deutsches Wikipedia Korpus

Erneut liegen die p-Werte deutlich über dem Signifikanzniveau. Eine Erklärung dafür könnte sein, dass (i) nur zwanzig Dokumente zu wenig für effizientes Topic Modeling sind und dementsprechend zu spezielle (bzw. verwirrende) Topics generiert werden, oder (ii) ‚literarische Topics‘ zu speziell für ein Evaluationskorpus wie Wikipedia sind, das vornehmlich (semi-)wissenschaftliche Texte enthält.

In der Umfrage<sup>9</sup> wurde kein einziges Topic als eindeutig sinnvoll oder sinnlos deklariert. Diese Uneinigkeit lässt zunächst darauf schließen, dass die Topics tatsächlich zu speziell sind bzw. bei den Probanden Unsicherheit auslöst.

#### 4.5.2 Röder Topics

Röder et al. (2015) stellen die Datensätze ihrer Experimente (Korpus, Topics, Goldstandard) zur Verfügung<sup>10</sup>. Um nach den bescheidenen Ergebnissen zuvor nun überhaupt die Validität der implementierten Klassen zu überprüfen, wurden die in Tabelle 2 zusammengefassten Experimente mit den Röder Daten durchgeführt und mit den Ergebnissen von Röder et al. verglichen:

	<b>Zeit</b>	<b>Korrelation von Röder</b>	<b>Spearman- Korrelation</b>	<b>p-Wert</b>
<b>UMass</b>	3 min 10 s	0.093	0.182	0.069
<b>UMass (nor- malisiert)</b>	3 min 9 s	-	-0.159	0.114
<b>UCI</b>	19 s	0.473	0.304	0.002
<b>UCI (norma- lisiert)</b>	18 s	0.438	-0.189	0.061

Tabelle 2: Röder Topics (5 × 100) und englisches Wikipedia Korpus

<sup>9</sup> Die ausgewertete Umfrage ist der beiliegenden CD zu entnehmen.

<sup>10</sup> <<http://139.18.2.164/mroeder/palmetto/datasets/>>

Mit  $\alpha = 0.1$  konnten die Ergebnisse von Röder et al. teilweise annähernd rekonstruiert werden. Für UMass wurde sogar eine höhere Korrelation nachgewiesen.

### 4.5.3 Wikipedia

Um nun zu überprüfen, ob ‚literarische Topics‘ für das Wikipedia Korpus zu speziell sind, wurde in 50.000 Iterationen ein Topic Model mit 255 zufällig ausgewählten Artikeln aus dem deutschen Wikipedia Korpus trainiert. Aus 100 Topics wurden 10 zufällig ausgewählt und vier Personen zur Bewertung vorgelegt. Die Ergebnisse der Kohärenzberechnungen sind Tabelle 4 zu entnehmen:

	<b>Zeit</b>	<b>Kohärenz (Median)</b>	<b>Spearman- Korrelation</b>	<b>p-Wert</b>
<b>UMass</b>	28 min 23 s	-0.067	0.865	0.001
<b>UMass (normalisiert)</b>	29 min 55 s	0.017	-0.693	0.026
<b>UCI</b>	9 min 44 s	-0.247	0.215	0.551
<b>UCI (normalisiert)</b>	12 min 42 s	0.056	-0.718	0.021

Tabelle 4: Wikipedia Topics (10 × 10) und deutsches Wikipedia Korpus

Bis auf UCI sind sämtliche Korrelationen signifikant. Besonders bemerkenswert ist das Ergebnis UMass, welches nahezu perfekt mit dem Goldstandard korreliert. Interessant ist hier, dass ein intrinsisches Maß bei der extrinsischen Evaluation einen besonders hohen Korrelationskoeffizienten erzielt und das extrinsische Maß keinen Zusammenhang herstellen kann.

## 5. Fazit und Ausblick

Im Rahmen der vorliegenden Arbeit wurden zwei Kohärenzmaße zur Evaluation von Topic Models vorgestellt (2.2) und implementiert (4.). Sämtliche Experimente mit einer Sammlung literarischer Texte, sowie der deutschen und englischen Wikipedia wurden dokumentiert und erläutert.

Nach zunächst bescheidenen Erfolgen (4.5.1) konnten teilweise die Ergebnisse von Röder et al. (2015) signifikant rekonstruiert werden (4.5.2). Vielversprechend waren die Ergebnisse der Experimente mit den Wikipedia Topics und der

deutschen Wikipedia als Evaluationskorpus (4.5.3). Hier konnte besonders bei UMass eine sehr hohe Korrelation mit dem Goldstandard nachgewiesen werden.

Das Modul bedarf einiger Verbesserungen und Ergänzungen, so sollte die Segmentierung der Topics ausgebaut werden. Außerdem könnte das *sliding window* Konzept, bei dem jedes Dokument im Korpus in  $n$  Teile aufgespalten wird und jeder Teil als neues Dokument behandelt wird, bei der Ermittlung der Häufigkeiten der Topic Keys zu besseren Ergebnissen führen. Trotzdem ist die Evaluation semantischer Kohärenz bei Topic Models kein gelöstes Problem. Weitere Fortschritte im Forschungsbereich der distributionellen Semantik können zur Lösung dieses Problems maßgeblich beitragen.

## 6. Literaturverzeichnis

- Nikolaos Aletras, Mark Stevenson**, Evaluating Topic Coherence Using Distributional Semantics, *Proceedings of the 10<sup>th</sup> International Conference on Computational Semantics* (2013), 13–22.
- David M. Blei, Andrew Y. Ng, Michael I. Jordan**, Latent Dirichlet Allocation, *Journal of Machine Learning Research* 3 (2003), 993–1022.
- Laurence BonJour**, *The Structure of Empirical Knowledge*, Cambridge (1985).
- Kenneth Ward Church, Patrick Hanks**, Word Association Norms, Mutual Information, and Lexicography, *Computational Linguistics* 16 (1990), 22–29.
- Zellig S. Harris**, Distributional Structure, *WORD* 10 (1954), 146–162.
- Christian Kracht**, *Der gelbe Bleistift*, Köln (2002).
- Andrew McCallum**, MALLET, A Machine Learning for Language Toolkit (2002), <<http://mallet.cs.umass.edu>>.
- David Mimno, Hanna M. Wallach, Edmund Talley, Miriam Leenders, Andrew McCallum**, Optimizing Semantic Coherence in Topic Models, *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (2011), 262–272.
- David Newman, Jey Han Lau, Karl Grieser, Timothy Baldwin**, Automatic Evaluation of Topic Coherence, *The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (2010), 100–108.

**Radim Řehůřek**, Gensim, Topic Modelling for Humans (2009),

<http://radimrehurek.com/gensim>.

**Michael Röder, Andreas Both, Alexander Hinneburg**, Exploring the Space of Topic Coherence Measures, *Proceedings of the 8<sup>th</sup> International Conference on Web Search and Data Mining* (2015).

**Keith Stevens, Philip Kegelmeyer, David Andrzejewski, David Buttler**, Exploring Topic Coherence Over Many Models and Many Topics, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (2012), 952–961.

## 7. Selbstständigkeitserklärung

Hiermit versichere ich, Severin Simmler, 2028090, dass ich die vorliegende Arbeit „Automatische Evaluation semantischer Kohärenz bei Topic Models“ selbständig verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und die Prüfungsleistung bisher oder gleichzeitig keiner anderen Prüfungsbehörde vorgelegt habe. Alle Stellen, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, habe ich einzeln durch Angaben der Quelle, auch der benutzten Sekundärliteratur, als Entlehnung kenntlich gemacht. Zusätzlich reiche ich die Arbeit auch in elektronischer Form als Datei beim Dozenten ein.

---

Ort, Datum

---

Unterschrift

## Was ist ein Plagiat?

Ein Plagiat bei einer schriftlichen Arbeit liegt vor, wenn bei einer schriftlichen Arbeit der Text oder Teile des Texts aus anderen Arbeiten (Büchern, Zeitschriften, dem Internet usw.) wörtlich oder sinngemäß ohne Angabe der Quelle übernommen oder übersetzt und damit fälschlicherweise als eigene geistige Leistung ausgegeben werden.