

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	07-202-1103A	Wahlpflicht

### Modultitel **Advanced Statistics**

**Modultitel (englisch)** Advanced Statistics

**Empfohlen für:** 1./3. Semester

**Verantwortlich** Professur für Statistik

**Dauer** 1 Semester

**Modulturnus** jedes Wintersemester

**Lehrformen**

- Vorlesung "Advanced Statistics" (2 SWS) = 30 h Präsenzzeit und 70 h Selbststudium = 100 h
- Übung "Advanced Statistics" (1 SWS) = 15 h Präsenzzeit und 35 h Selbststudium = 50 h

**Arbeitsaufwand** 5 LP = 150 Arbeitsstunden (Workload)

**Verwendbarkeit** • M.Sc. Data Science

**Ziele**

Die Studierenden können die theoretischen Grundlagen - d.h. mehrdimensionale Wahrscheinlichkeitsverteilungen, Inferenzkonzepte und deren Anwendung auf Regressionsmodelle - sowie den aktuellen Stand der Forschung im Bereich der Wahrscheinlichkeitsrechnung und induktiven Statistik darstellen und beschreiben. Sie sind in der Lage, selbständig geeignete, dem aktuellen Forschungsstand entsprechende Methoden und Strategien zur Lösung statistischer Problemstellungen auszuwählen, anzuwenden sowie die ermittelten Ergebnisse zu interpretieren und kritisch zu beurteilen.

Die Studierenden können in Übungen die Charakteristika komplexer statistischer Problemstellungen erkennen, passende Modelle zu deren Modellierung auswählen und mittels geeigneter Schätzmethoden anpassen. Hierbei sind sie in der Lage, das Softwarepakete R als Werkzeug für statistische Analysen einzusetzen, die Programmierung einfacher Routinen durchzuführen und ihre erzielten Ergebnisse kritisch auszuwerten.

**Inhalt** Mehrdimensionale Wahrscheinlichkeitsverteilungen, Inferenzkonzepte, Maximum-Likelihood-Schätzung und Likelihood-basierte Testverfahren, Generalisierte Lineare Regressionsmodelle

**Teilnahmevoraussetzungen** keine

**Literaturangabe** Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten** Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Klausur 80 Min., mit Wichtung: 1</b>	
	Vorlesung "Advanced Statistics" (2SWS)
	Übung "Advanced Statistics" (1SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-201-2502	Wahlpflicht

### Modultitel **Praktikum der IT-Sicherheit**

**Modultitel (englisch)** IT-Security Lab

**Empfohlen für:** 1./2./3. Semester

**Verantwortlich** Professur für Data Privacy and Security

**Dauer** 1 Semester

**Modulturnus** unregelmäßig

**Lehrformen** • Praktikum "IT-Sicherheit" (2 SWS) = 30 h Präsenzzeit und 120 h Selbststudium = 150 h

**Arbeitsaufwand** 5 LP = 150 Arbeitsstunden (Workload)

**Verwendbarkeit** • M.Sc. Data Science  
• M.Sc. Informatik

**Ziele** Nach der aktiven Teilnahme am Modul sind die Studierenden in der Lage, Angriffe auf IT-Systeme aktiv zu verstehen, Gegenmaßnahmen zu erarbeiten und durchzusetzen.

Die Studierenden sind anschließend imstande, Sicherheitslücken selbstständig zu identifizieren, Angriffe zu implementieren und Sicherheitstechniken zu entwickeln.

Sie lernen arbeitsteilig und effizient im Team zu arbeiten, Probleme zu identifizieren und zielgerichtet zu lösen.

**Inhalt** Das Praktikum ist zweigeteilt. Im ersten Teil werden die Studierenden schrittweise an grundlegende Techniken und Werkzeuge herangeführt, mit einem direkten Anschluss an die im Modul „Grundlagen der IT-Sicherheit“ vermittelten Konzepte. Dabei kommen insbesondere Analysewerkzeuge für Programme und Netzwerkverkehr zum Einsatz, sowie typische Werkzeuge für Penetrationstests. Im zweiten Teil werden die Werkzeuge und Techniken auf konkrete Sicherheitsobjekte angewendet. Dies beinhaltet auch das Programmieren von automatisierten Abläufen mit dem Ziel, Teilnetze nach Schwachstellen zu durchsuchen oder umfangreiche Netzwerk-Mitschnitte mit Verfahren aus dem Machine Learning zu analysieren. Die Studierenden organisieren sich dabei selbstständig in Teams, und erarbeiten konkrete Zielstellungen, Analyse- und Ausführungspläne. Die Studierenden dokumentieren dabei ihr Vorgehen und präsentieren ihre Zwischenstände und Ergebnisse.

Das Praktikum richtet sich an Studierende, die entweder bereits Kenntnisse der IT-Sicherheit mitbringen oder dies im Rahmen der Vorlesung „Grundlagen der IT-Sicherheit“ parallel erwerben möchten.

Inhaltlich setzt das Praktikum detaillierte Kenntnisse und persönliches Interesse im Bereich Programmierung, Netzwerkprotokolle und Rechnernetze voraus, da beispielsweise Netzwerksniffer ohne diese Kenntnisse nicht sinnvoll eingesetzt werden können. Ein paralleler Erwerb dieser Kenntnisse ist didaktisch nicht sinnvoll.

Ein Gespräch mit der Abteilung wird im Vorfeld zur Anmeldung empfohlen.

**Teilnahmevoraussetzungen**

vertiefte Kenntnisse im Bereich Programmierung (Python, Java, C oder C++), Netzwerkprotokolle und Rechnernetze sowie Grundkenntnisse in Machine Learning und in der IT-Sicherheit  
Eine zeitgleiche Belegung des Moduls 10-201-2503 wird empfohlen.

**Literaturangabe**

Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Praktikumsleistung (Präsentation (20 Min.) mit schriftlicher Ausarbeitung (4 Wochen)), mit Wichtung: 1</b>	
	Praktikum "IT-Sicherheit" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-201-2503	Wahlpflicht

<b>Modultitel</b>	<b>Grundlagen der IT-Sicherheit</b>
<b>Modultitel (englisch)</b>	Introduction to IT-Security
<b>Empfohlen für:</b>	1./2./3. Semester
<b>Verantwortlich</b>	Professur für Data Privacy and Security
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	unregelmäßig
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• Vorlesung "Grundlagen der IT-Sicherheit" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h</li> <li>• Übung "Grundlagen der IT-Sicherheit" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h</li> </ul>
<b>Arbeitsaufwand</b>	5 LP = 150 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	<ul style="list-style-type: none"> <li>• M.Sc. Data Science</li> <li>• M.Sc. Informatik</li> </ul>
<b>Ziele</b>	<p>Nach der aktiven Teilnahme am Modul sind die Studierenden in der Lage</p> <ul style="list-style-type: none"> <li>- grundlegende Begriffe zu definieren, deren Spezifika zu benennen und zwischen ihnen zu unterscheiden</li> <li>- ausgewählte Bedrohungen der Computersicherheit zu erläutern und auf deren Gegenmaßnahmen einzugehen</li> <li>- Probleme der Netzwerksicherheit und dem Internet zu benennen</li> <li>- verbreitete Bewertungskriterien sicherer Software und System zu erläutern</li> <li>- ausgewählte Sicherheitsmodelle zu benennen und deren Eignung zu beurteilen</li> <li>- algorithmische Lösungsansätze zu erläutern und anzuwenden</li> <li>- im Rahmen des Security Engineering verschiedene Analysen zu erarbeiten</li> </ul>
<b>Inhalt</b>	<p>Vorlesung "Grundlagen der IT-Sicherheit"</p> <ul style="list-style-type: none"> <li>- Herangehensweisen, Modell, Konzepte und Verfahren der IT-Sicherheit</li> <li>- Bedrohungen und Angriffstechniken, z.B. Computerviren oder Pufferüberläufe</li> <li>- soziotechnische Bedrohungen wie Spam und Phishing</li> <li>- Gegenmaßnahmen wie Malwarescanner und Sandboxing</li> <li>- Internet- und Netzwerksicherheit</li> <li>- Herangehensweisen beim Security Engineering, z.B. BSI-Sicherheitsprozess</li> <li>- Sicherheitsmodelle, Standards und Zertifizierungen</li> <li>- Bewertungskriterien sicherer Software und Computersysteme</li> </ul> <p>Übung "Grundlagen der IT-Sicherheit"</p> <ul style="list-style-type: none"> <li>- Absicherung von Computersystemen</li> <li>- Kryptographie und deren Anwendungen</li> <li>- Security Engineering</li> </ul>
<b>Teilnahmevoraussetzungen</b>	keine
<b>Literaturangabe</b>	Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben.  
Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Klausur 60 Min., mit Wichtung: 1</b>	
	Vorlesung "Grundlagen der IT-Sicherheit" (2SWS)
	Übung "Grundlagen der IT-Sicherheit" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2131	Wahlpflicht

<b>Modultitel</b>	<b>Einführung in Soziale Netzwerke</b> Kernmodul
<b>Modultitel (englisch)</b>	Introduction to Social Networks Key Module
<b>Empfohlen für:</b>	1./3. Semester
<b>Verantwortlich</b>	Professur für Rechnernetze und Verteilte Systeme
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	jedes Wintersemester
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• Vorlesung "Soziale Netzwerke" (2 SWS) = 30 h Präsenzzeit und 75 h Selbststudium = 105 h</li> <li>• Übung "Soziale Netzwerke" (1 SWS) = 15 h Präsenzzeit und 30 h Selbststudium = 45 h</li> </ul>
<b>Arbeitsaufwand</b>	5 LP = 150 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	<ul style="list-style-type: none"> <li>• M.Sc. Data Science</li> <li>• M.Sc. Digital Humanities</li> <li>• Kernmodul der Angewandten Informatik im M.Sc. Informatik</li> <li>• Kernmodul der Praktischen Informatik im M.Sc. Informatik</li> <li>• M.Sc. Wirtschaftsinformatik</li> </ul>
<b>Ziele</b>	<p>Nach der aktiven Teilnahme am Modul Soziale Netzwerke sind die Studierenden in der Lage,</p> <ul style="list-style-type: none"> <li>- die mathematische Zusammenhänge und Verfahren sozialer Netzwerke sowie deren technischen Herausforderungen umfassend und anschaulich mündlich zu erläutern.</li> <li>- das Verständnis für mathematische Zusammenhänge und Verfahren sozialer Netzwerke zu erarbeiten, so dass Fragestellungen und Rechenaufgaben zu diesen mathematischen Zusammenhängen und Verfahren schriftlich bearbeitet bzw. gelöst werden können und</li> <li>- in kleinen Gruppen Fragestellungen zu bearbeiten und zu diskutieren.</li> </ul>
<b>Inhalt</b>	<ul style="list-style-type: none"> <li>- Grundlagen Sozialer Netzwerke</li> <li>- Cold Start Problem und Viralität</li> <li>- Inhalts- und Strukturbasiert Empfehlungsverfahren</li> <li>- Konsumenten und Communities in sozialen Netzwerken</li> <li>- Metriken und Bewertung von Empfehlungsverfahren</li> <li>- Promotion Kampagnen und Spam-Erkennung</li> <li>- Angriffsresistente Empfehlungssysteme</li> <li>- Fallbeispiele: twitter, Instagram, YouTube, u.a.</li> </ul>
<b>Teilnahmevoraussetzungen</b>	Teilnahme am Modul "Rechnernetze" (10-201-2107), Modul "Internetanwendungen" (10-201-2106), Vertiefungsmodul "Rechnernetze und Internetanwendungen" (10-201-2102) oder gleichwertige Kenntnisse
<b>Literaturangabe</b>	Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben.  
Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Mündliche Prüfung 20 Min., mit Wichtung: 1</b>	
<i>Prüfungsvorleistung: Klausur (15 Min.)</i>	
	Vorlesung "Soziale Netzwerke" (2SWS)
	Übung "Soziale Netzwerke" (1SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2132	Wahlpflicht

<b>Modultitel</b>	<b>Soziale Netzwerke</b>
	Vertiefungsmodul
<b>Modultitel (englisch)</b>	Social Networks
	In-Depth Module
<b>Empfohlen für:</b>	1./3. Semester
<b>Verantwortlich</b>	Professur für Rechnernetze und Verteilte Systeme
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	jedes Wintersemester
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• Vorlesung "Soziale Netzwerke" (2 SWS) = 30 h Präsenzzeit und 75 h Selbststudium = 105 h</li> <li>• Übung "Soziale Netzwerke" (1 SWS) = 15 h Präsenzzeit und 30 h Selbststudium = 45 h</li> <li>• Seminar "Soziale Netzwerke" (2 SWS) = 30 h Präsenzzeit und 120 h Selbststudium = 150 h</li> </ul>
<b>Arbeitsaufwand</b>	10 LP = 300 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	<ul style="list-style-type: none"> <li>• M.Sc. Data Science: Ergänzungsbereich</li> <li>• M.Sc. Digital Humanities</li> <li>• M.Sc. Informatik: Vertiefungsmodul</li> </ul>
<b>Ziele</b>	<p>Nach der aktiven Teilnahme am Modul Soziale Netzwerke sind die Studierenden in der Lage,</p> <ul style="list-style-type: none"> <li>- die mathematische Zusammenhänge und Verfahren sozialer Netzwerke sowie deren technischen Herausforderungen umfassend und anschaulich mündlich zu erläutern,</li> <li>- das Verständnis für mathematische Zusammenhänge und Verfahren sozialer Netzwerke zu erarbeiten, so dass Fragestellungen und Rechenaufgaben zu diesen mathematischen Zusammenhängen und Verfahren schriftlich bearbeitet bzw. gelöst werden können, und</li> <li>- die Ergebnisse wissenschaftlicher Veröffentlichungen kritisch zu bewerten, mit anderen Veröffentlichungen zu vergleichen und mit Studierenden zu diskutieren.</li> </ul>
<b>Inhalt</b>	<ul style="list-style-type: none"> <li>- Grundlagen Sozialer Netzwerke</li> <li>- Cold Start Problem und Viralität</li> <li>- Inhalts- und Strukturbasiert Empfehlungsverfahren</li> <li>- Konsumenten und Communities in sozialen Netzwerken</li> <li>- Metriken und Bewertung von Empfehlungsverfahren</li> <li>- Promotion Kampagnen und Spam-Erkennung</li> <li>- Angriffsresistente Empfehlungssysteme</li> <li>- Fallbeispiele: twitter, Instagram, YouTube, u.a.</li> </ul>
<b>Teilnahmevoraussetzungen</b>	Ein erfolgreich abgeschlossenes Modul zur Wahrscheinlichkeitstheorie und Statistik, wie z.B. das Modul "Wahrscheinlichkeitstheorie" (10-201-1802) an der Universität Leipzig oder gleichwertige Kenntnisse.

**Literaturangabe** Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten** Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung:</b>	
Mündliche Prüfung 20 Min., mit Wichtung: 1 <i>Prüfungsvorleistung: (schriftliche Klausur (15 Min.))</i>	Vorlesung "Soziale Netzwerke" (2SWS)
Referat (20 Min.) mit schriftlicher Ausarbeitung, mit Wichtung: 1	Übung "Soziale Netzwerke" (1SWS)
	Seminar "Soziale Netzwerke" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2133	Wahlpflicht

### Modultitel **Künstliche Neuronale Netze, Deep Learning, Maschinelles Lernen und Signalverarbeitung**

Vertiefungsmodul

**Modultitel (englisch)** Artificial Neural Networks, Deep Learning, Machine Learning and Signal Processing

In-Depth Module

**Empfohlen für:** 1./3. Semester

**Verantwortlich** Professur für Neuromorphe Informationsverarbeitung

**Dauer** 1 Semester

**Modulturnus** jedes Wintersemester

**Lehrformen**

- Vorlesung "Künstliche neuronale Netze und Maschinelles Lernen" (2 SWS) = 30 h Präsenzzeit und 70 h Selbststudium = 100 h
- Vorlesung "Signalverarbeitung und Deep Learning" (2 SWS) = 30 h Präsenzzeit und 70 h Selbststudium = 100 h
- Praktikum "KI" (2 SWS) = 30 h Präsenzzeit und 70 h Selbststudium = 100 h

**Arbeitsaufwand** 10 LP = 300 Arbeitsstunden (Workload)

**Verwendbarkeit**

- Informatikmodul im M.Sc. Bioinformatik
- M.Sc. Data Science
- Vertiefungsmodul Technische Informatik im M.Sc. Informatik
- M.Sc. Medizininformatik

**Ziele**

Nach der aktiven Teilnahme am Modul sind die Studierenden in der Lage

- grundlegende Begriffe aus den beiden Vorlesungen zu definieren und zu erklären
- ausgewählte Verfahren und Algorithmen zu beschreiben und zu analysieren
- algorithmische Lösungsansätze zu erklären und diese selbstständig auf Problemstellungen anzuwenden
- Aufgabenstellung praktisch in Form eines Software-basierten Verfahrens zu lösen.

**Inhalt**

Die Studierenden sollen die grundlegenden überwachten und unüberwachten Lernverfahren und Algorithmen der Künstlichen Neuronalen Netze, des Deep Learnings und des Maschinellen Lernens sowie der Signalverarbeitung verstehen und die wesentlichen Lösungsansätze auf Problemstellung der industriellen und wissenschaftlichen Anwendungen anwenden können.

Als Praktikumsleistung stehen auf vielfachen Wunsch der Studierenden eine Projektarbeit in Gruppenarbeit bzw. 5 Versuche in Einzelarbeit zur Wahl. Die individuelle Festlegung hierzu erfolgt vor Beginn des Praktikums oder zu Beginn der Veranstaltung.

**Teilnahmevoraussetzungen** Nicht für Studierende, die bereits am Kernmodul "Künstliche Neuronale Netze und Maschinelles Lernen" 10-202-2128 teilgenommen haben.

**Literaturangabe** Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben.  
Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung:</b>	
Mündliche Prüfung 25 Min., mit Wichtung: 1	Vorlesung "Künstliche neuronale Netze und Maschinelles Lernen" (2SWS)
<i>Prüfungsvorleistung: (Praktikumsleistung (Bearbeitungsdauer 10 Wochen) mit Abschlusspräsentation (15 Minuten))</i>	
	Vorlesung "Signalverarbeitung und Deep Learning" (2SWS)
	Praktikum "KI" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2137	Wahlpflicht

<b>Modultitel</b>	<b>KI und Ethik</b> Seminar modul
<b>Modultitel (englisch)</b>	AI and Ethics Seminar Module
<b>Empfohlen für:</b>	1./3. Semester
<b>Verantwortlich</b>	Professur für Neuromorphe Informationsverarbeitung
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	jedes Wintersemester
<b>Lehrformen</b>	• Seminar "AI and Ethics" (2 SWS) = 30 h Präsenzzeit und 120 h Selbststudium = 150 h
<b>Arbeitsaufwand</b>	5 LP = 150 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	• M.Sc. Informatik: Seminar modul • M.Sc. Bioinformatik • M.Sc. Medizininformatik • M.Sc. Data Science: Ergänzungsbereich
<b>Ziele</b>	Nach der aktiven Teilnahme am Modul sind die Studierenden in der Lage - Einen wissenschaftlichen Text zur KI und Ethik zu analysieren, in eigenen Worten darstellen, - Inhalte aus solchen Texten geeignet aufbereitet darstellen - Fragestellung zu bearbeiten und zu diskutieren, - Eine wissenschaftliche Veröffentlichung zu verfassen, die den formellen Anforderungen einer Konferenz entspricht.
<b>Inhalt</b>	Selbständige Bearbeitung einer aktuellen Forschungsarbeit zur KI und Ethik und Vortrag darüber.  - Lehrsprache: English oder Deutsch - Prüfungssprache: English oder Deutsch Die Festlegung hierzu erfolgt vor der Moduleinschreibung auf elektronischem Weg (Vorlesungsverzeichnis) oder zu Beginn der Veranstaltung durch den Dozenten/die Dozentin.
<b>Teilnahmevoraussetzungen</b>	Kenntnisse im Bereich maschinelles Lernen / künstliche neuronale Netze sind empfehlenswert, aber nicht zwingend erforderlich
<b>Literaturangabe</b>	Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.
<b>Vergabe von Leistungspunkten</b>	Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Referat (25 Min.) mit schriftlicher Ausarbeitung (8 Wochen), mit Wichtung: 1</b>	
	Seminar "AI and Ethics" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2201	Wahlpflicht

<b>Modultitel</b>	<b>Wissenschaftliche Visualisierung</b> Vertiefungsmodul
<b>Modultitel (englisch)</b>	Scientific Visualization In-Depth Module
<b>Empfohlen für:</b>	1./3. Semester
<b>Verantwortlich</b>	Professur für Bild- und Signalverarbeitung
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	jedes Wintersemester
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• Vorlesung "Wissenschaftliche Visualisierung" (4 SWS) = 60 h Präsenzzeit und 120 h Selbststudium = 180 h</li> <li>• Praktikum "Wissenschaftliche Visualisierung" (4 SWS) = 60 h Präsenzzeit und 60 h Selbststudium = 120 h</li> </ul>
<b>Arbeitsaufwand</b>	10 LP = 300 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	<ul style="list-style-type: none"> <li>• M.Sc. Bioinformatik</li> <li>• M.Sc. Biologie</li> <li>• M.Sc. Data Science</li> <li>• M.Sc. Informatik</li> <li>• M.Sc. Medizininformatik</li> <li>• M.Sc. Wirtschaftspädagogik (zweites Fach Informatik)</li> <li>• Staatsexamen Lehramt Informatik</li> </ul>
<b>Ziele</b>	Nach der aktiven Teilnahme am Vertiefungsmodul Wissenschaftliche Visualisierung können die Studierenden alle Grundkonzepte der wissenschaftlichen Visualisierung skizzieren. Die Studierenden können beurteilen, welches wissenschaftliche Visualisierungsverfahren der für eine bestimmte Aufgabe am besten geeignete Ansatz ist. Die Studierenden können grundlegende Verfahren der wissenschaftlichen Visualisierung in Programmen selbständig implementieren.
<b>Inhalt</b>	<p>Das Modul umfasst eine Vorlesung und ein Praktikum, die beide zu belegen sind. Die wissenschaftliche Visualisierung beschäftigt sich mit der Nutzung der Computergrafik zur Generierung von Bildern und Animationen, die einer verbesserten Auswertung von Experimenten und Simulationen durch den Menschen dienen. Sie gehört in vielen Disziplinen zu den grundlegenden Techniken der Datenauswertung. Die Vorlesung behandelt vor allem Prinzipien, Methoden und erfolgreiche Beispiele zur Visualisierung von Felddaten, wie sie bei Simulationen und Messungen in Physik, Chemie, Meteorologie und den Ingenieurwissenschaften, aber auch der Medizin auftreten. Ferner werden Aspekte des Entwurfs von Visualisierungssystemen behandelt. Themen sind u. a. Datenpräsentation, Grundlagen aus Theorie und Anwendungsdomänen, direkte Visualisierung, geometrische Visualisierung, Direct Volume Rendering, topologische Visualisierung, struktur- und merkmalsorientierte Visualisierung, Visualisierungssysteme.</p> <p>Das Praktikum dient der eigenständigen Umsetzung von Verfahren aus der Vorlesung und aktuellen Publikationen, wobei auch Erfahrungen zur Entwicklung ganzer Visualisierungssysteme gewonnen werden.</p>

**Teilnahmevoraussetzungen** keine

**Literaturangabe** Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten** Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Mündliche Prüfung 30 Min., mit Wichtung: 1</b>	
<i>Prüfungsvorleistung: Praktikumsleistung (Präsentation (30 Min) mit schriftlicher Ausarbeitung) im Praktikum, Bearbeitungszeit (8 Wochen)</i>	
	Vorlesung "Wissenschaftliche Visualisierung" (4SWS)
	Praktikum "Wissenschaftliche Visualisierung" (4SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2205	Wahlpflicht

### Modultitel **Graphen und biologische Netze**

**Modultitel (englisch)** Graphs and Biological Nets

**Empfohlen für:** 1./3. Semester

**Verantwortlich** Professur für Bioinformatik

**Dauer** 1 Semester

**Modulturnus** jedes Wintersemester

**Lehrformen**

- Vorlesung "Einführungsvorlesung Graphentheorie" (2 SWS) = 30 h Präsenzzeit und 56 h Selbststudium = 86 h
- Vorlesung "Aktuelle Forschungsthemen aus dem Bereich Graphen und biologische Netze" (1 SWS) = 15 h Präsenzzeit und 28 h Selbststudium = 43 h
- Seminar "Seminar zur Spezialvorlesung" (1 SWS) = 15 h Präsenzzeit und 28 h Selbststudium = 43 h
- Praktikum "Praktikum" (3 SWS) = 45 h Präsenzzeit und 83 h Selbststudium = 128 h

**Arbeitsaufwand** 10 LP = 300 Arbeitsstunden (Workload)

**Verwendbarkeit**

- M.Sc. Biochemie
- M.Sc. Bioinformatik
- M.Sc. Biologie
- M.Sc. Data Science
- M.Sc. Medizininformatik

**Ziele**

Nach der aktiven Teilnahme am Modul "Graphen und Biologische Netze" sind die Studierenden in der Lage:

- grundlegende Begriffe und Konzepte der Graphentheorie zu formulieren und zu erklären,
- biologische Fragestellungen als graphentheoretische Probleme zu modellieren und mithilfe geeigneter algorithmischer Ansätze zu lösen und
- die Ergebnisse im Kontext der biologischen Fragestellung zu interpretieren und kritisch zu diskutieren.

**Inhalt**

Grundvorlesung:

- Grundlegende Eigenschaften von Graphen: Zusammenhang, Planarität, Kreise, Färbungen
- Zufallsgraphen

Spezialvorlesung/ Seminar: aktuelle Forschungsthemen, z.B.

- Metabolische Netzwerke: Flussanalyse, Organisationen, Netzwerk-Evolution
- Genregulationsnetzwerke: Dynamik, Stabilität,
- Modelle komplexer biologischer Netzwerke: Wachsende Netzwerke, Skalenfreiheit, Selbstähnlichkeit

- Lehrsprache: englisch oder deutsch
- Prüfungssprache: englisch oder deutsch

Die Festlegung hierzu erfolgt vor der Moduleinschreibung auf elektronischem Weg

(Vorlesungsverzeichnis) oder zu Beginn der Veranstaltung durch den Dozenten/ die Dozentin.

**Teilnahmevoraussetzungen** keine

**Literaturangabe** Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten** Für die Vergabe von Leistungspunkten müssen alle vorgesehenen Studienleistungen erbracht sowie die Prüfungsleistung bestanden sein.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Mündliche Prüfung 30 Min., mit Wichtung: 1</b>	
<i>Prüfungsvorleistung: • Referat (30 Min.) im Seminar,</i>	
<i>• Praktikumsleistung als schriftliche Ausarbeitung im Praktikum, Bearbeitungszeit 8 Wochen</i>	
	Vorlesung "Einführungsvorlesung Graphentheorie" (2SWS)
	Vorlesung "Aktuelle Forschungsthemen aus dem Bereich Graphen und biologische Netze" (1SWS)
	Seminar "Seminar zur Spezialvorlesung" (1SWS)
	Praktikum "Praktikum" (3SWS)

\* Diese Prüfungsleistungen müssen bestanden sein.

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2207	Wahlpflicht

### Modultitel **Sequenzanalyse und Genomik**

**Modultitel (englisch)** Sequence Analysis and Genomics

**Empfohlen für:** 1./3. Semester

**Verantwortlich** Professur für Bioinformatik

**Dauer** 1 Semester

**Modulturnus** jedes Wintersemester

**Lehrformen**

- Vorlesung "Einführungsvorlesung Sequenzanalyse und Genomik" (2 SWS) = 30 h Präsenzzeit und 56 h Selbststudium = 86 h
- Vorlesung "Spezialvorlesung Sequenzanalyse und Genomik" (1 SWS) = 15 h Präsenzzeit und 28 h Selbststudium = 43 h
- Übung "Sequenzanalyse und Genomik" (1 SWS) = 15 h Präsenzzeit und 28 h Selbststudium = 43 h
- Praktikum "Sequenzanalyse und Genomik" (4 SWS) = 60 h Präsenzzeit und 68 h Selbststudium = 128 h

**Arbeitsaufwand** 10 LP = 300 Arbeitsstunden (Workload)

**Verwendbarkeit**

- Wahlpflichtmodul im M.Sc. Biochemie
- M.Sc. Bioinformatik
- Wahlpflichtmodul im M.Sc. Biologie
- M.Sc. Data Science
- Vertiefungsmodul im M.Sc. Informatik
- M.Sc. Medizininformatik
- M.Sc. Wirtschaftspädagogik (zweites Fach Informatik)
- Lehramt Informatik

**Ziele**

Nach der aktiven Teilnahme am Modul "Bioinformatik von RNA und Proteinstrukturen" sind die Studierenden in der Lage

- Sequenzdaten im biologischen Kontext zu interpretieren,
- die grundlegenden Algorithmen zum Sequenzvergleich in hinreichender Tiefe zu verstehen, um die geeigneten Werkzeuge für konkrete Anwendungen auszuwählen,
- die grundlegenden Algorithmen zum Sequenzvergleich anzuwenden und in einfacher Weise zu modifizieren,
- einfache Aufgabenstellungen aus der vergleichenden Genomik eigenständig zu bearbeiten und
- die Ergebnisse der praktischen Arbeit zu präsentieren und kritisch zu diskutieren.

**Inhalt**

Vorlesung "Sequenzanalyse und Genomik":

- Exakte und approximative Suche in Sequenzdaten
- lokale und globale Alignierung von Sequenzen
- Phylogenetische Rekonstruktion in Theorie und Praxis

Eine Spezialvorlesung wird auf einem der folgenden Themengebiete angeboten:

- "Evolutionäre Algorithmen": Kombinatorische Optimierungs-Probleme; Simulated Annealing; Werte-Landschaften; Genetische Algorithmen; Genetic Programming.

- "Hidden-Markov-Modelle in der Bioinformatik": Grundlagen von HMMs: Baum-Welch- und Viterbi-Algorithmus; Parameterschätzung; paarweise Alignments mit HMMs; Profile-HMMs für Sequenzfamilien; multiple Alignments mit Lernen von Profile-HMMs.
- "Präbiotische Evolution": Astrophysikalische Grundlagen; Präbiotische Chemie; Chemische Reaktionsnetzwerke; Die RNA Welt und alternative Szenarien; Mathematische Modelle: Quasispecies, Hyperzyklus, und Co.; Der Genetische Code
- "Populationsgenetik": Einführung in die theoretischen Grundlagen und die Werkzeuge der Populationsgenetik sowie der Populationsgenomik.
- "Epigenetik": Arten der epigenetischen Modifikationen; Begriffsdefinition Epigenetik; Einführung in die experimentellen Techniken mit Schwerpunkt auf ihre Auswertung; Mapping von Sequenzierungsdaten; Peak-Calling Verfahren;
- "Algorithmen für Hochdurchsatzsequenzierung": Hochgeschwindigkeitsalignentalgorithmen basierend auf Suffix Arrays und der Burrows Wheeler Transformation wie "BWA", "BOWTIE" oder "segemehl". Algorithmen zur Rekonstruktion von Genomen basierend auf De Bruijn Graphen oder String Graphen.

Ein Praktikum wird auf einem der folgenden Themengebiete angeboten:

- "Nukleinsäuren": Praxisnaher Umgang mit Standard-Programmen (u.a. "blast", "clustal" und "muscle") zur genomweiten Suche und zum Sequenzvergleich; Suche nach strukturierter Information, wie z.B. proteinkodierenden Regionen, nicht-kodierenden RNAs oder regulatorischen Elementen in Genomen unter Zuhilfenahme aktueller Werkzeuge und Methoden (z.B. "Proteinortho", "RNAz" oder "Augustus"); Umgang mit Datenquellen wie dem "UCSC Genome Browser".
- "Phylogenetische Rekonstruktion": Rekonstruktion von Phylogenien mit Standard-Werkzeugen wie "phylip", "MEGA" oder "NeighborNet"; problemgerechte Auswahl einer Methode (Maximum Parsimony, Maximum Likelihood oder distanzbasiert); visuelle Darstellung von Ereignissen und Veränderungen auf evolutionären Zeitskalen (u.a. mit "TreeView" oder "iTOL")
- "Epigenetik": Einführung in grundlegende Auswertungsprogramme wie "bedtools" oder "UCSCtools" sowie Programme zur Erstauswertung von Sequenzierungsexperimenten wie "cutadapt", "fastqc" oder "segemehl".
- "Populationsgenetik": Verfahren zur Analyse von Daten zu genetischen Polymorphismen und genomischer Diversität sowie mathematischer Modelle zur Modellierung von populationsgenetischen Effekten (wie Mutation, Drift und Selektion) in der Evolution.

In den Praktika wird zum selbstständigen Arbeiten angeleitet, nicht die Ergebnisse, sondern das Erstellen und Verfeinern von Lösungsansätzen stehen im Vordergrund.

Eine Übung begleitet die Vorlesung, in der vorgestellte Algorithmen implementiert und vertieft und vorgestellte Programme angewandt werden.

**Teilnahmevoraussetzungen**

keine

**Literaturangabe**

Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Mündliche Prüfung 30 Min., mit Wichtung: 1</b>	
<i>Prüfungsvorleistung: Praktikumsbericht im Praktikum, Bearbeitungszeit 8 Wochen</i>	
	Vorlesung "Einführungsvorlesung Sequenzanalyse und Genomik" (2SWS)
	Vorlesung "Spezialvorlesung Sequenzanalyse und Genomik" (1SWS)
	Übung "Sequenzanalyse und Genomik" (1SWS)
	Praktikum "Sequenzanalyse und Genomik" (4SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2303	Wahlpflicht

<b>Modultitel</b>	<b>Logik in der Informatik</b>
<b>Modultitel (englisch)</b>	Logic in Computer Science
<b>Empfohlen für:</b>	1./3. Semester
<b>Verantwortlich</b>	Professur für Grundlagen der Wissensrepräsentation
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	jedes Wintersemester
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• Vorlesung "Logik in der Informatik" (3 SWS) = 45 h Präsenzzeit und 60 h Selbststudium = 105 h</li> <li>• Übung "Logik in der Informatik" (1 SWS) = 15 h Präsenzzeit und 30 h Selbststudium = 45 h</li> </ul>
<b>Arbeitsaufwand</b>	5 LP = 150 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	<ul style="list-style-type: none"> <li>• M.Sc. Data Science: Ergänzungsbereich</li> <li>• M.Sc. Informatik: Kernmodul der Theoretischen Informatik</li> </ul>
<b>Ziele</b>	Nach aktiver Teilnahme an diesem Modul sind die Studierenden vertraut mit den grundlegenden Konzepten und Resultaten der Logik in der Informatik. Sie können die Ausdrucksstärke und Komplexität einer Logik einschätzen und analysieren und für eine gegebene Anwendung eine geeignete Logik auswählen. Zudem sind sie in der Lage, typische Algorithmen für logische Schlussfolgerungsprobleme auszuführen, zu implementieren und an neue Logiken anzupassen.
<b>Inhalt</b>	<p>In diesem Modul werden Vorlesungen zu folgenden Themen angeboten:</p> <ul style="list-style-type: none"> <li>- Fortgeschrittene Themen der Logik</li> <li>Endliche Modelltheorie, Spiele, Sätze von Büchi, Fagin und Courcelle, 0/1-Gesetze</li> <li>- Datenbanktheorie</li> <li>Konjunktive Anfragen, Azyklizität, Beschränkte Baumweite, Datalog, probabilistische Datenbanken, Mappings und Views</li> <li>- Beschreibungslogik</li> <li>Ontologien, Subsumption, Komplexität, ALC und EL, Bisimulationen</li> </ul> <p>Die Übungen sind vorlesungsbegleitend und ermöglichen eine eigenständige und vertiefende Beschäftigung mit dem Vorlesungsmaterial.</p>
<b>Teilnahmevoraussetzungen</b>	Teilnahme am Modul "Logik" (10-201-2108-1) oder gleichwertige Kenntnisse
<b>Literaturangabe</b>	Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.
<b>Vergabe von Leistungspunkten</b>	Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen****Modulprüfung: Klausur 90 Min., mit Wichtung: 1***Prüfungsvorleistung: Erwerb eines studienbegleitenden Übungsscheines (6 Übungsblätter mit Hausaufgaben, 1 Woche Bearbeitungszeit pro Übungsschein)*

Vorlesung "Logik in der Informatik" (3SWS)

Übung "Logik in der Informatik" (1SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2314	Wahlpflicht

### Modultitel **Information Retrieval**

**Modultitel (englisch)** Information Retrieval

**Empfohlen für:** 1./2./3./4. Semester

**Verantwortlich** Juniorprofessur für Text Mining

**Dauer** 1 Semester

**Modulturnus** unregelmäßig

**Lehrformen**

- Vorlesung "Information Retrieval" (2 SWS) = 30 h Präsenzzeit und 70 h Selbststudium = 100 h
- Übung "Information Retrieval" (1 SWS) = 15 h Präsenzzeit und 70 h Selbststudium = 85 h
- Praktikum "Information Retrieval" (3 SWS) = 45 h Präsenzzeit und 70 h Selbststudium = 115 h

**Arbeitsaufwand** 10 LP = 300 Arbeitsstunden (Workload)

**Verwendbarkeit**

- M.Sc. Data Science
- M.Sc. Digital Humanities
- Vertiefungsmodul im M.Sc. Informatik

**Ziele**

Die Suche nach Informationen, die dazu beitragen, eine Wissenslücke zu schließen oder die Lösung einer komplexen Aufgabe voran zu treiben, ist ein alltäglicher Vorgang. Informationssysteme, die die Suche in digitalen Daten ermöglichen, werden als Suchmaschinen bezeichnet und assistieren beim Auffinden (engl. "Retrieval") von Informationen. Anders als beim Datenretrieval ist die Suche typischerweise von vagen Anfragen und unsicherem sowie unvollständigem Wissen gekennzeichnet. Die Rolle von Suchmaschinen beim Wissenstransfer von Produzenten zu Konsumenten von Informationen ist Gegenstand der Forschung im Information Retrieval. In der Vorlesung werden grundlegende und fortgeschrittene Konzepte, Methoden und der mathematische Hintergrund des Information Retrieval zur Entwicklung von Suchmaschinen für unstrukturierte Textdaten vermittelt.

Nach der aktiven Teilnahme am Modul sind die Studierenden in der Lage, die Retrieval-Probleme realer Suchdomänen zu identifizieren, die Konzepte und Methoden des Information Retrieval zu definieren und anzuwenden, eine Suchmaschine für eine gegebene Suchdomäne zu entwickeln, die Qualität einer Suchmaschine systematisch zu evaluieren, wohlinformierte Entscheidungen über den Ansatz verschiedener Retrievalmodelle zu treffen und praktische Gesichtspunkte für die Verbesserung von Suchsystemen analysieren und einschätzen zu können. Die Studierenden sind weiterhin in der Lage, selbständig aktuelle Ansätze aus der Forschung zu reproduzieren und unter ausreichender Supervision auch eigene Suchverfahren zu entwickeln.

**Inhalt**

In der Vorlesung werden grundlegende und fortgeschrittene Konzepte und Methoden des Information Retrieval sowie die entsprechenden mathematischen Hintergründe vermittelt. Dazu gehören die Architektur von Suchmaschinen, die Akquise, Vorverarbeitung und Informationsextraktion aus unstrukturierten Textdaten, Algorithmen und Datenstrukturen für Indexe und Anfrageverarbeitung,

grundlegende Retrievalmodelle und Evaluierungsverfahren, Learning-to-Rank-Algorithmen, Query Understanding, Neuronale Retrieval-Modelle, Retrieval-Axiome, und Online-Evaluierungsverfahren, sowie ausgewählte Themen aus der aktuellen Forschung.

- Lehrsprache: englisch oder deutsch
- Prüfungssprache: englisch oder deutsch

Die Festlegung hierzu erfolgt vor der Moduleinschreibung auf elektronischem Weg (Vorlesungsverzeichnis) oder zu Beginn der Veranstaltung durch den Dozenten/die Dozentin.

**Teilnahmevoraussetzungen**

keine

**Literaturangabe**

- C.D. Manning, P. Raghavan, H. Schütze. Introduction to Information Retrieval.
- W.B. Croft, D. Metzler, T. Strohman. Search Engines: Information Retrieval in Practice.

Weitere Hinweise zu Literaturangaben erfolgen in der Lehrveranstaltung.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung:</b>	
Klausur 60 Min., mit Wichtung: 2	Vorlesung "Information Retrieval" (2SWS)
	Übung "Information Retrieval" (1SWS)
Projektarbeit: Präsentation (30 Min.) mit schriftlicher Ausarbeitung (8 Wochen), mit Wichtung: 1	Praktikum "Information Retrieval" (3SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2316	Wahlpflicht

### Modultitel **Applied Language Technologies**

**Modultitel (englisch)** Applied Language Technologies

**Empfohlen für:** 1./2./3./4. Semester

**Verantwortlich** Juniorprofessur für Text Mining und Retrieval

**Dauer** 1 Semester

**Modulturnus** unregelmäßig

**Lehrformen**

- Seminar "Big Data und Language Technologies" (2 SWS) = 30 h Präsenzzeit und 90 h Selbststudium = 120 h
- Praktikum "Big Data und Language Technologies" (4 SWS) = 60 h Präsenzzeit und 120 h Selbststudium = 180 h

**Arbeitsaufwand** 10 LP = 300 Arbeitsstunden (Workload)

**Verwendbarkeit**

- M.Sc. Data Science: Bereich Datenanalyse
- M.Sc. Digital Humanities
- M.Sc. Informatik: Vertiefungsmodul

**Ziele** Nach der aktiven Teilnahme am Modul sind die Studierenden je nach gewähltem Seminarthema in der Lage, skalierbare Algorithmen des NLP und des IR auf großen Datenmengen anzuwenden, Methoden des Maschinellen Lernens und insbesondere des Deep Learnings anzuwenden, ein algorithmisches Forschungsproblem zu erfassen und in Wort und Schrift zu erklären.

**Inhalt** Als Sprachtechnologien werden Methoden und Werkzeuge zur Analyse, Modifikation und Generierung menschlicher Sprache bezeichnet. Sie werden mit erforscht und entwickelt, um zwischenmenschliche Interaktionen sowie Interaktionen zwischen Mensch und Maschine in natürlicher Sprache zu unterstützen. Sprachtechnologien sind Grundlage zahlreicher intelligenter Anwendungen wie Suchmaschinen, Übersetzungssysteme, Dialog- und Konversationssystemen oder Argumentationssystemen und viele mehr. Erforscht werden sie in den Bereichen Natural Language Processing (NLP), Information Retrieval (IR).

Eine wichtige Grundlage der Sprachtechnologien bilden Verfahren der Künstlichen Intelligenz (KI), des Maschinellen Lernens (ML) und insbesondere des Deep Learning (DL). Der Einsatz letzterer erfordert einerseits skalierbare verteilte Architekturen sowie Sprachdaten in großem Umfang, die als Trainingsbeispiele zur Modellbildung ausgewertet werden. Eine der primären Datenquellen hierfür bildet das Web.

Die Webis-Forschungsgruppe betreibt einen großen, modernen Hochleistungsrechner mit umfangreicher Speicher- und Rechenkapazität sowie aktuellen Middlewares (<https://webis.de/facilities.html>). Dort wird ein Webausschnitt im Umfang von Petabytes für Forschungszwecke vorgehalten, der die Gegenwart und Vergangenheit des Webs widerspiegelt – ein einmaliger Forschungsgegenstand.

Die Studierenden erhalten eine anwendungsorientierte Ausbildung in NLP, IR, Big Data und Deep Learning, lösen Aufgaben und untersuchen eigenverantwortlich interessante Forschungsfragen. Dank der Größe des vorhandenen Clusters und den Kompetenzen der Webis-Gruppe in den Bereichen NLP, IR, und Big Data, bietet dieses Seminar ein außergewöhnliches Ausbildungsniveau

- Lehrsprache: englisch oder deutsch
- Prüfungssprache: englisch oder deutsch

Die Festlegung hierzu erfolgt vor der Moduleinschreibung auf elektronischem Weg (Vorlesungsverzeichnis) oder zu Beginn der Veranstaltung durch den Dozenten/die Dozentin.

**Teilnahmevoraussetzungen**

Teilnahme an den Modulen "Grundlagen des maschinellen Lernens" (10-201-2315) und "Linguistische Informatik" (10-201-2317) oder gleichwertige Kenntnisse

**Literaturangabe**

Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Projektarbeit: schriftliche Ausarbeitung (8 Wochen) und Präsentation (30 Min.), mit Wichtung: 1</b>	
	Seminar "Big Data und Language Technologies" (2SWS)
	Praktikum "Big Data und Language Technologies" (4SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2317	Wahlpflicht

### Modultitel **Advanced Language Technologies**

**Modultitel (englisch)** Advanced Language Technologies

**Empfohlen für:** 1./2./3./4. Semester

**Verantwortlich** Juniorprofessur für Text Mining und Retrieval

**Dauer** 1 Semester

**Modulturnus** unregelmäßig

**Lehrformen**

- Vorlesung "Advanced Language Technologies" (2 SWS) = 30 h Präsenzzeit und 70 h Selbststudium = 100 h
- Übung "Advanced Language Technologies" (1 SWS) = 15 h Präsenzzeit und 70 h Selbststudium = 85 h
- Praktikum "Advanced Language Technologies" (3 SWS) = 45 h Präsenzzeit und 70 h Selbststudium = 115 h

**Arbeitsaufwand** 10 LP = 300 Arbeitsstunden (Workload)

**Verwendbarkeit**

- M.Sc. Data Science: Bereich Datenanalyse
- M.Sc. Digital Humanities
- M.Sc. Informatik: Vertiefungsmodul

**Ziele**

Nach der aktiven Teilnahme am Modul sind die Studierenden in der Lage, Probleme des Natural Language Processing zu identifizieren, Konzepte und Methoden des Natural Language Processing zu definieren und anzuwenden, ein Verfahren zur Sprachverarbeitung für ein gegebenes Problem zu entwickeln, die Qualität eines Ansatzes zur Sprachverarbeitung zu evaluieren, wohlinformierte Entscheidungen über den Ansatz verschiedener Methoden der Sprachverarbeitung zu treffen und praktische Gesichtspunkte für die Verbesserung von System zur Sprachverarbeitung analysieren und einschätzen zu können. Geeignete Algorithmen und Verfahren des Maschinellen Lernens und der Künstlichen Intelligenz wurden erarbeitet und im Kontext der Sprachverarbeitung zum Einsatz gebracht. Studierende sind weiterhin in der Lage, selbständig aktuelle Ansätze aus der Forschung zu reproduzieren und unter ausreichender Supervision auch eigene Verfahren zu entwickeln.

**Inhalt**

Natürliche Sprachen sind von Menschen gesprochene Sprachen, die sich dynamisch und weitgehend ungesteuert entwickelt haben. Die Verarbeitung natürlicher Sprache (engl. "Natural Language Processing", kurz NLP) zählt zu den zentralen Herausforderungen der Informatik im Bereich der angewandten Künstlichen Intelligenz. Ziele sind unter anderem computergestützt Menschen beim Schreiben zu unterstützen, Texte zu identifizieren die eine gesprochene oder geschriebene Frage zu beantworten, Texte automatisch einzusortieren, ihnen spezifische Informationen zu entnehmen, sie zusammenzufassen oder zu übersetzen. Damit sich Maschinen nahtlos in eine von und für Menschen gemachte Umgebung einfügen können, sollen natürliche Sprachen als Benutzerschnittstelle dienen. Letztlich möchte man sich mit einer Maschine unterhalten können, als wäre sie ein Mensch. All die daraus resultierenden Technologien werden auch Sprachtechnologien (engl. "Language Technologies") genannt.

Im Modul werden fortgeschrittene Konzepte und Methoden der Sprachtechnologien sowie die entsprechenden mathematischen Hintergründe vermittelt und praktisch erprobt. Dazu gehören eine Auswahl fortgeschrittener Wort-, Syntax-, Semantik- und Pragmatik-Phänomene, die Modellierung von Text mit Methoden des Maschinellen Lernens, sowie fortgeschrittene Anwendungen wie die automatische Übersetzung, Paraphrasierung und Zusammenfassung von Texten, die Autorschaftsanalyse, das Argumentationsmining, die Informationsextraktion, das Question Answering und Konversations- und Dialogsysteme bis hin zu aktuellen Forschungsthemen, wie beispielsweise das Sprachverstehen (engl. "Language Understanding") oder kausale Sprachen sowie kausale Inferenz (engl. "Causal Language" und "Causal Inference") als zukünftige Herausforderungen der Künstlichen Intelligenz.

- Lehrsprache: englisch oder deutsch  
 - Prüfungssprache: englisch oder deutsch  
 Die Festlegung hierzu erfolgt vor der Moduleinschreibung auf elektronischem Weg (Vorlesungsverzeichnis) oder zu Beginn der Veranstaltung durch den Dozenten/die Dozentin.

**Teilnahmevoraussetzungen** "Grundlagen des Maschinellen Lernens" (10-201-2315) oder gleichwertige Kenntnisse

**Literaturangabe** - D. Jurafsky, J.H. Martin. Speech and Language Processing.  
 - C.D. Manning, H. Schütze. Foundations of Natural Language Processing.

Weitere Hinweise zu relevanter Literatur erfolgen in der Lehrveranstaltung.

**Vergabe von Leistungspunkten** Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung:</b>	
Klausur 60 Min., mit Wichtung: 2	Vorlesung "Advanced Language Technologies" (2SWS)
	Übung "Advanced Language Technologies" (1SWS)
Projektarbeit: schriftliche Ausarbeitung (8 Wochen) und Präsentation (30 Min.), mit Wichtung: 1	Praktikum "Advanced Language Technologies" (3SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2342	Wahlpflicht

<b>Modultitel</b>	<b>Linguistic Annotation and Data Extraction with XQuery</b>
<b>Modultitel (englisch)</b>	Linguistic Annotation and Data Extraction with XQuery
<b>Empfohlen für:</b>	1./3. Semester
<b>Verantwortlich</b>	Leitung des Instituts für Informatik
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	unregelmäßig
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• Vorlesung "Linguistic annotation and data extraction with XQuery" (2 SWS) = 30 h Präsenzzeit und 120 h Selbststudium = 150 h</li> <li>• Praktikum "Linguistic annotation and data extraction with XQuery" (0 SWS) = 0 h Präsenzzeit und 150 h Selbststudium = 150 h</li> </ul>
<b>Arbeitsaufwand</b>	10 LP = 300 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	<ul style="list-style-type: none"> <li>• M.Sc. Data Science</li> <li>• M.Sc. Digital Humanities</li> <li>• Ergänzungsfach im M.Sc. Informatik</li> </ul>
<b>Ziele</b>	<ol style="list-style-type: none"> <li>1) Die Studierenden werden mit den Grundlagen der Linguistik vertraut gemacht, die notwendig sind, um linguistische Annotationen im informatischen Kontext umfassend verstehen und anwenden zu können.</li> <li>2) Vermittlung von XQuery zur Anwendung und Abfrage von Dokumenten und Annotationen</li> </ol>
<b>Inhalt</b>	<p>Vorlesung:</p> <ol style="list-style-type: none"> <li>1) Die verschiedenen Layers der Linguistischen Annotation: Morphologie, Syntax und Semantik</li> <li>2) XQuery: Navigieren eines XML Dokuments</li> <li>3) XQuery: FLWOR expression</li> <li>4) XQuery: Funktionen</li> </ol> <p>Praktikum:</p> <ol style="list-style-type: none"> <li>1) Entwicklung von Ressourcen und Annotieren eines Pilotkorpus` für eine bestimmte Sprache</li> </ol> <p>Die Lehrveranstaltungen werden in englischer Sprache gehalten. Studien- und Prüfungsleistungen sind in englischer Sprache zu erbringen.</p>
<b>Teilnahmevoraussetzungen</b>	keine
<b>Literaturangabe</b>	Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.
<b>Vergabe von Leistungspunkten</b>	Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Praktikumsleistung (Programmierung eines Skripts) 90 Min., mit Wichtung: 1</b>	
	Vorlesung "Linguistic annotation and data extraction with XQuery" (2SWS)
	Praktikum "Linguistic annotation and data extraction with XQuery" (0SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2345	Wahlpflicht

### Modultitel **Software Engineering für KI-Systeme**

Vertiefungsmodul

**Modultitel (englisch)** Software Engineering for AI-enabled Systems

In-Depth Module

**Empfohlen für:** 1./3. Semester

**Verantwortlich** Professur für Softwaresysteme

**Dauer** 1 Semester

**Modulturnus** jedes Wintersemester

**Lehrformen**

- Vorlesung "Software Engineering für KI-Systeme" (4 SWS) = 60 h Präsenzzeit und 70 h Selbststudium = 130 h
- Übung "Software Engineering für KI-Systeme" (1 SWS) = 15 h Präsenzzeit und 45 h Selbststudium = 60 h
- Projekt "Software Engineering für KI-Systeme" (2 SWS) = 30 h Präsenzzeit und 80 h Selbststudium = 110 h

**Arbeitsaufwand** 10 LP = 300 Arbeitsstunden (Workload)

**Verwendbarkeit**

- Vertiefungsmodul M.Sc. Informatik
- Wahlpflichtmodul M.Sc. Data Science
- Ergänzungsbereich

**Ziele**

Nach erfolgreichem Abschluss des Kurses sollten Studierende in der Lage sein

- Tradeoffs beim Entwerfen von Produktionssystemen mit KI-Komponenten zu analysieren, verschiedene Qualitätseigenschaften zusätzlich zur Genauigkeit, wie z.B. Operationskosten, Latenz, Updatebarkeit, Erklärbarkeit zu analysieren;
- Produktionssysteme, die robust gegenüber Fehlern von KI-Komponenten sind, zu implementieren;
- Fehlertolerante und skalierbare Daten-intensive Infrastrukturen zum Lernen, Ausliefern, Versionieren und Experimentieren von Modellen zu entwerfen;
- Qualität der gesamten Machine-Learning Pipeline durch Testautomatisierung und anderen Qualitätssichernden Techniken, einschließlich automatisierten Checks, Data Drifts, Feedback-Schleifen, sicher zu stellen;
- System zu bauen, die in Produktion getestet werden können und Deployment-Pipelines bauen, die vorsichtige Rollouts und Canary Testing ermöglichen
- Fairness, Ethik und Moral sowie Sicherheit beim Bauen von komplexen KI-basierten Systemen zu berücksichtigen
- Effektiv im Team bestehend aus Softwaretechnikern und Data Analysten zu kommunizieren.

After taking this course, among others, students should be able to

- analyze tradeoffs for designing production systems with AI-components, analyzing various qualities beyond accuracy such as operating cost, latency, updateability, and explainability
- implement production-quality systems that are robust to mistakes of AI components
- design fault-tolerant and scalable data infrastructure for learning models, serving models, versioning, and experimentation

- ensure quality of the entire machine learning pipeline with test automation and other quality assurance techniques, including automated checks for data quality, data drift, feedback loops, and model quality
- build systems that can be tested in production and build deployment pipelines that allow careful rollouts and canary testing
- consider fairness, and security when building complex AI-enabled systems
- communicate effectively in teams with both software engineers and data analysts

## Inhalt

Der Kurs vermittelt Wissen zum Management, Anforderungsanalyse und Bau von Softwaresystemen, die eine signifikante Machine-Learning bzw. KI-Komponente beinhalten. Der Kurs diskutiert, wie man eine Idee zusammen mit einem KI-Modell (z.B. in Form von Skripten oder Jupyter Notebooks) aufgreift und dies als Teil eines skalierbaren, wartbaren Systems (z.B. App, Web Anwendung, IoT Anwendung) ausliefert. Anstelle des Fokus auf die Modellierung oder die KI Algorithmen zu legen, nimmt der Kurs eine Arbeitsbeziehung mit einem Data Scientists an und fokussiert auf das Design, Implementierung, Betrieb und Gewährleistung des umliegenden Softwaresystems. Insbesondere sind Inhalte hierbei Projekt- und Teammanagement, Anforderungsanalyse, Architektur von SW/ML Systemen, Experimente, MLOps, Testing und Deployment von KI-Systemen, Validität und Fairness.

Der Kurs zielt auf Softwaretechniker und -entwickler ab, die verstehen wollen, welche spezifischen Herausforderung beim Arbeiten mit KI-Komponenten existieren und an Datenwissenschaftlern, die verstehen wollen, wie man von einem Prototyp bzw. einfachen KI Modell in die Produktion geht. Der Kurs hilft somit bei der Kommunikation und Zusammenarbeit beider Rollen.

- Lehrsprache: englisch oder deutsch
- Prüfungssprache: englisch oder deutsch

Die Festlegung hierzu erfolgt vor der Moduleinschreibung auf elektronischem Weg (Vorlesungsverzeichnis) oder zu Beginn der Veranstaltung durch den Dozenten/die Dozentin.

The course takes a software engineering perspective on managing, requirements engineering, and building software systems with a significant machine learning or AI component. It discusses how to take an idea and a model developed by a data scientist (e.g., scripts and Jupyter notebook) and deploy it as part of scalable and maintainable system (e.g., mobile apps, web applications, IoT devices). Rather than focusing on modeling and learning itself, this course assumes a working relationship with a data scientist and focuses on issues of design, implementation, operation, and assurance and how those interact with the data scientist's modeling. The content includes project and team management, requirements engineering for AI systems, architecture for SW systems, experimentation, MLOps, testing and deployment of AI-systems, validity and fairness.

This course is aimed at software engineers who want to understand the specific challenges of working with AI components and at data scientists who want to understand the challenges of getting a prototype model into production; it facilitates communication and collaboration between both roles.

## Teilnahmevoraussetzungen

keine

## Literaturangabe

Building Intelligent Systems (Geof Hulton), Apress, ISBN-10: 1484234316.  
 Building Machine Learning Powered Applications von Emmanuel Ameisen (ISBN: 9781492045113)  
 ML Ops: Operationalizing Data Science von David Sweenor, Steven Hillion, Dan Rope, Dev Kannabiran, Thomas Hill, Michael O'Connell (ISBN: 9781492074656)  
 Managing Data Science von Kirill Dubovikov (ISBN: 9781838826321)  
 Accelerated DevOps with AI, ML & RPA: Non-Programmer's Guide to AIOps & MLOps von Stephen Fleming (ISBN: 978-1702763653)  
 Agile-AI von Carlo Appugliese, Paco Nathan, William S. Roberts (ISBN: 9781492074977)  
 Practical Fairness von Aileen Nielsen (ISBN: 9781492075738)  
 Machine Learning Logistics von Ted Dunning & Ellen Friedman (ISBN: 978-1-491-

99759-8)

Accelerate: The Science of Lean Software and DevOps: Building and Scaling High Performing Technology Organizations von Nicole Forsgren und Gene Kim und Jez Humble (ISBN: 978-1942788331)

Weitere Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Klausur 120 Min., mit Wichtung: 1</b>	
<i>Prüfungsvorleistung: Projektarbeit zu den Themen MLOps, Deployment und Feedbackloop), Bearbeitungszeit: 14 Wochen</i>	
	Vorlesung "Software Engineering für KI-Systeme" (4SWS)
	Übung "Software Engineering für KI-Systeme" (1SWS)
	Projekt "Software Engineering für KI-Systeme" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-DIH-0001	Wahlpflicht

<b>Modultitel</b>	<b>Verfahren und Anwendungen in den Digital Humanities</b>
<b>Modultitel (englisch)</b>	Methods and Applications in the Digital Humanities
<b>Empfohlen für:</b>	1./3. Semester
<b>Verantwortlich</b>	Juniorprofessur für Computational Humanities
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	jedes Wintersemester
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• Vorlesung "Einführung in grundlegende Verfahren und Anwendungen der Digital Humanities" (2 SWS) = 30 h Präsenzzeit und 60 h Selbststudium = 90 h</li> <li>• Übung "Praktischer Einsatz von Verfahren und Anwendungen in den Digital Humanities" (2 SWS) = 30 h Präsenzzeit und 60 h Selbststudium = 90 h</li> <li>• Praktikum "Projektarbeit" (2 SWS) = 30 h Präsenzzeit und 90 h Selbststudium = 120 h</li> </ul>
<b>Arbeitsaufwand</b>	10 LP = 300 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	<ul style="list-style-type: none"> <li>• M.Sc. Data Science</li> <li>• M.Sc. Digital Humanities</li> <li>• M.Sc. Journalismus</li> </ul>
<b>Ziele</b>	<p>Nach der aktiven Teilnahme am Modul "Verfahren und Anwendungen in den Digital Humanities" sind die Studierenden in der Lage:</p> <ol style="list-style-type: none"> <li>1) grundlegende Forschungsfelder und Forschungsansätze sowie auch aktuelle Trends und Herausforderungen in den Digital Humanities zu skizzieren,</li> <li>2) konkrete Methoden und Tools zur Durchführung von Digital Humanities-Forschungsprojekten korrekt anzuwenden,</li> <li>3) zu entscheiden, welches Verfahren der für eine Aufgabe am besten geeignete Ansatz ist.</li> </ol>
<b>Inhalt</b>	<p>Das Modul umfasst eine Vorlesung zur Einführung in grundlegende Verfahren und Anwendungen in den Digital Humanities. Diese Verfahren und Anwendungen werden in einer begleitenden Übung praktisch erprobt und in einem Praktikum von den Studierenden in Projektarbeiten eigenständig angewandt.</p> <p>Vorlesung:</p> <ul style="list-style-type: none"> <li>- Theoretische Grundlagen der Digital Humanities</li> <li>- Kurzer geschichtlicher Überblick zu Methoden und Anwendungen in den Digital Humanities</li> <li>- Aktuelle Trends und Herausforderungen in den Digital Humanities</li> <li>- Typische Tätigkeiten und Arbeitsabläufe in einem Digital Humanities-Projekt</li> <li>- Systematisierung bestehender Methoden und Anwendungen</li> <li>- Vorstellen konkreter Methoden aus unterschiedlichen Anwendungsbereichen (Text, Bild, Musik, etc.) der Digital Humanities.</li> </ul> <p>Übung:</p> <ul style="list-style-type: none"> <li>- Aufbau einer grundlegenden "Digital Humanities"-Toolbox und praktische Anwendung der Verfahren und Anwendungen aus der Vorlesung</li> </ul>

<b>Teilnahmevoraussetzungen</b>	keine
<b>Literaturangabe</b>	Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.
<b>Vergabe von Leistungspunkten</b>	Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

### Prüfungsleistungen und -vorleistungen

<b>Modulprüfung: Projektarbeit: schriftliche Ausarbeitung (8 Wochen) und Präsentation (30 Min.), mit Wichtung: 1</b>	
	Vorlesung "Einführung in grundlegende Verfahren und Anwendungen der Digital Humanities" (2SWS)
	Übung "Praktischer Einsatz von Verfahren und Anwendungen in den Digital Humanities" (2SWS)
	Praktikum "Projektarbeit" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-INF-BI01	Wahlpflicht

<b>Modultitel</b>	<b>Statistisches Lernen</b> Vertiefungsmodul
<b>Modultitel (englisch)</b>	Statistical Learning In-Depth Module
<b>Empfohlen für:</b>	1./3. Semester
<b>Verantwortlich</b>	Professur für Integrative Bioinformatik
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	jedes Wintersemester
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• E-Learning-Veranstaltung "Grundlagen des statistischen Lernens" (0 SWS) = 0 h Präsenzzeit und 60 h Selbststudium = 60 h</li> <li>• Seminar "Grundlagen des statistischen Lernens" (2 SWS) = 30 h Präsenzzeit und 120 h Selbststudium = 150 h</li> <li>• Übung "Statistisches Lernen mit R" (2 SWS) = 30 h Präsenzzeit und 60 h Selbststudium = 90 h</li> </ul>
<b>Arbeitsaufwand</b>	10 LP = 300 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	<ul style="list-style-type: none"> <li>• M.Sc. Bioinformatik</li> <li>• M.Sc. Data Science</li> <li>• M.Sc. Digital Humanities</li> <li>• M.Sc. Informatik</li> </ul>
<b>Ziele</b>	<p>Nach der aktiven Teilnahme am Modul "Statistisches Lernen" sind die Studierenden in der Lage:</p> <ul style="list-style-type: none"> <li>- grundlegende Verfahren der Statistik korrekt anzuwenden,</li> <li>- verschiedene Verfahren des Maschinellen Lernens zu erklären, zu vergleichen, und zu komplexen Workflows zu verbinden und</li> <li>- Workflows der (bio)informatischen Datenanalyse in der Statistiksprache R zu implementieren.</li> </ul>
<b>Inhalt</b>	<p>Seminar und Praktikum "Grundlagen des statistischen Lernens"</p> <ul style="list-style-type: none"> <li>- Wahrscheinlichkeitsbegriff, deskriptive Statistik, Verteilungen, statistisches Testen</li> <li>- Statisches Lernen, lineare Regression, Klassifikation</li> <li>- Resampling-Methoden, Modellwahl, Regularisierung</li> <li>- Supervised und unsupervised (machine) learning, Dimensionsreduktion</li> <li>- Explorative Datenanalyse</li> <li>- Hochdimensionale systembiologische Daten, multiples Testen</li> <li>- Einführung in die reproduzierbare Datenanalyse und Programmieren in R anhand von Beispieldatensätzen</li> <li>- Storytelling with data</li> </ul>
<b>Teilnahmevoraussetzungen</b>	Grundkenntnisse in Statistik oder Biometrie oder gleichwertige Kenntnisse
<b>Literaturangabe</b>	Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben.  
Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Klausur 60 Min., mit Wichtung: 1</b>	
<i>Prüfungsvorleistung: Praktikumsleistung (10 elektronische Testate (Bearbeitungsdauer je 3 Tage))</i>	
	E-Learning-Veranstaltung "Grundlagen des statistischen Lernens" (0SWS)
	Seminar "Grundlagen des statistischen Lernens" (2SWS)
	Übung "Statistisches Lernen mit R" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-INF-DS01	Pflicht

### Modultitel **Skalierbare Datenbanktechnologien 1**

**Modultitel (englisch)** Scalable Database Technologies 1

**Empfohlen für:** 1. Semester

**Verantwortlich** Professur für Informatik (Datenbanken)

**Dauer** 1 Semester

**Modulturnus** jedes Wintersemester

**Lehrformen**

- Vorlesung "Skalierbare Datenbanktechnologien 1" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h
- Vorlesung "Skalierbare Datenbanktechnologien 2" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h
- Seminar "Skalierbare Datenbanktechnologien" (2 SWS) = 30 h Präsenzzeit und 120 h Selbststudium = 150 h

**Arbeitsaufwand** 10 LP = 300 Arbeitsstunden (Workload)

**Verwendbarkeit** • M.Sc. Data Science

**Ziele** Nach der aktiven Teilnahme am Modul "Skalierbare Datenbanktechnologien 1" sind die Studierenden in der Lage verschiedene verteilte und parallele Datenmanagementsysteme anzugeben.

Sie können Eigenschaften und Architekturen von Integrationssystemen sowie Techniken zur Anfragebearbeitung und -optimierung in verteilten und parallelen Datenbanksystemen erklären.

Die Studierenden sind in der Lage wissenschaftliche Publikationen aus dem Bereich moderne Datenbanktechnologien, Cloud und Big Data zu erläutern und zu präsentieren.

Sie vermögen aktuelle Datenbanktechnologien selbständig in einer schriftlichen Ausarbeitung zu beurteilen.

Die Studierenden sind in der Lage, verschiedene Ansätze in der Gruppe zu diskutieren.

**Inhalt** Es werden mindestens zwei der folgenden Vorlesungen angeboten. Zwei Vorlesungen und das Seminar werden ausgewählt.

Vorlesung Mehrrechner-Datenbanksysteme

- Klassifikation von Mehrrechner-DBS
- Architektur von Verteilten DBS
- Datenverteilung
- Verteilte und parallele Anfrageoptimierung
- Transaktionsverwaltung in Verteilten DBS
- Replizierte DBS
- Cluster-DBS (Shared Disk).

Vorlesung Cloud Data Management

- Cloud Computing, Infrastrukturen und Dienste
- Verteilte Dateisysteme
- MapReduce Konzept, MapReduce im Kontext von Datenbanken
- Anwendungsimplementierung in verteilten Umgebungen und
- Optimierungstechniken
- Large-scale Datenanalyse, Analyse-Frameworks

Vorlesung Data Mining

- Skalierbare Algorithmen zur Analyse von großen Datenmengen
- Arbeit mit hochdimensionalen Daten
- Analyse von Datenströmen
- Netzwerkanalyse
- Assoziationsregeln, Clusteranalyse, Empfehlungsdienste
- Large-Scale Machine Learning

Vorlesung Implementierung von Datenbanksystemen I

- Aufbau von DBS (Schichtenmodell)
- Externspeicherverwaltung: Dateiverwaltung, Einsatz von Speicherhierarchien, Disk-Arrays, nicht-flüchtige Halbleiterspeicher
- Pufferverwaltung: Lokalität, Speicherallokation, Seitenersetzung
- Satzverwaltung: Satzadressierung, lange Felder, Column stores
- Indexstrukturen für DBS: B-Bäume, Hash-Verfahren, Grid-File, Text-Indizes, etc.
- Implementierung relationaler Operationen: Selektion, Joins, Sortierung
- Anfragebearbeitung: Übersetzung/Interpretation, Query-Optimierung.

Seminar Forschungsseminar Datenbanken

- Präsentieren und Diskussion von Arbeiten aus dem Gebiet der Datenbanktechnologie oder verwandten Gebieten
- Die Themenstellung richtet sich nach den aktuellen Entwicklungen auf dem Gebiet der Datenbanktechnologie bzw. verwandten Gebieten.
- Im Rahmen des Seminars ist eine Ausarbeitung zu einem Teilthema anzufertigen und über ihren Inhalt vorzutragen.

**Teilnahmevoraussetzungen**

keine

**Literaturangabe**

Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Klausur 120 Min., mit Wichtung: 2</b>	
	Vorlesung "Skalierbare Datenbanktechnologien 1" (2SWS)
	Vorlesung "Skalierbare Datenbanktechnologien 2" (2SWS)
Referat (60 Min.) mit schriftlicher Ausarbeitung (4 Wochen), mit Wichtung: 1	Seminar "Skalierbare Datenbanktechnologien" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-INF-DS201	Wahlpflicht

### Modultitel **Mathematische Grundlagen der Datenanalyse**

**Modultitel (englisch)** Mathematical Foundations of Data Analysis

**Empfohlen für:** 1. Semester

**Verantwortlich** Humboldt Professur für Künstliche Intelligenz

**Dauer** 1 Semester

**Modulturnus** jedes Wintersemester

**Lehrformen**

- Vorlesung "Mathematische Grundlagen der Datenanalyse" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h
- Seminar "Mathematische Grundlagen der Datenanalyse" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h

**Arbeitsaufwand** 5 LP = 150 Arbeitsstunden (Workload)

**Verwendbarkeit** • M.Sc. Data Science: Bereich Datenanalyse

**Ziele**

Nach der aktiven Teilnahme am sind die Studierenden in der Lage

- die Definitionen der Standardbegriffe der Datenwissenschaft und die damit verbundenen mathematischen Begriffe zu verstehen
- die mathematischen Beweise für die Funktionsweise häufig verwendeter Techniken zu verstehen
- die Algorithmen zu implementieren und Beispiele mit einem Computerprogramm (Julia) zu berechnen
- eine wissenschaftliche Präsentation zu halten

After actively taking part in the course, the students should understand

- The definitions of the standard tools in Data science, and understand the connections to the underlying mathematical ideas.
- Understand the mathematical proofs of how commonly used techniques work.
- Implement the algorithms and compute examples with a computer program.
- Present material beyond the class material through a project.

**Inhalt**

Lineare Algebra

- Unterräume
- Orthogonalität
- die Pseudo-inverse
- die Singluarwertzerlegung

Wahrscheinlichkeitstheorie

Thematische Schwerpunkte

- Netzwerkanalyse: Graphen und Laplace Matrix, das Spektrum eines Graphen, Markov-prozesse in Netzwerken, und Zentralitätsmaße. (2)
- Maschinelles Lernen: Daten, Modelle, Lernen, Regression im Statistischen Modell, Principal Component Analysis (Methode zur Dimensionsreduktion), und Support Vector Machines (Methode zur binären Klassifikation)
- Topologisch Datenanalyse: Simpliziale Komplexe und Homologie
- Matrizen und Tensoren: Matrizen von niedrigem Rang und Tensoren

Das Modul wird in englischer Sprache gehalten. Lehr- und Prüfungssprache ist

englisch.

Linear Algebra

- Subspaces
- Orthogonality
- The Pseudoinverse
- Singular value decomposition

Probability Theory

Afterwards we have 4 Thematic topics

- Network Analysis: graphs and the Laplace matrix, the spectrum of a graph, Markov processes in networks, centrality measures.
- Machine Learning: data, models, learning, regression in statistical models, principle component analysis (methods of dimension reduction), and support vector machines (methods of binary classification)
- Topological Data Analysis: simplicial complexes and homology
- Matrices and Tensors: matrices of low rank and tensors

**Teilnahmevoraussetzungen**

keine

**Literaturangabe**

Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Klausur 60 Min., mit Wichtung: 1</b>	
<i>Prüfungsvorleistung: Projektarbeit (8 Wochen) mit schriftlicher Ausarbeitung</i>	
	Vorlesung "Mathematische Grundlagen der Datenanalyse" (2SWS)
	Seminar "Mathematische Grundlagen der Datenanalyse" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	07-202-2302	Wahlpflicht

<b>Modultitel</b>	<b>Multivariate Datenanalyse und Data Mining</b>
<b>Modultitel (englisch)</b>	Multivariate Statistics and Data Mining
<b>Empfohlen für:</b>	2. Semester
<b>Verantwortlich</b>	Professur für Statistik
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	jedes Sommersemester
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• E-Learning-Veranstaltung "Multivariate Datenanalyse und Data Mining" (4 SWS) = 60 h Präsenzzeit und 140 h Selbststudium = 200 h</li> <li>• Übung "Multivariate Datenanalyse und Data Mining" (2 SWS) = 30 h Präsenzzeit und 70 h Selbststudium = 100 h</li> </ul>
<b>Arbeitsaufwand</b>	10 LP = 300 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	<ul style="list-style-type: none"> <li>• M.Sc. Data Science</li> <li>• M.Sc. Volkswirtschaftslehre (Economics)</li> </ul>
<b>Ziele</b>	<p>Die Studierenden können die theoretischen Grundlagen, d.h. Verfahren zur Dimensionsreduktion, Clusteranalyse, Diskriminanzanalyse, Klassifikations- und Assoziationsverfahren und künstliche neuronale Netze sowie den aktuellen Stand der Forschung im Bereich der Multivariaten Statistik sowie des Data Mining darstellen und beschreiben.</p> <p>Sie sind in der Lage, selbständig geeignete Methoden zur Lösung volks- und betriebswirtschaftlicher Problemstellungen auszuwählen, anzuwenden sowie die ermittelten Ergebnisse zu interpretieren und kritisch zu bewerten. Hierbei erweitern sie ihren Wissensstand neben statistischen Ansätzen um aus der Informatik entlehnte Methoden des Maschinenlernens.</p> <p>Die Studierenden können in Übungen die Charakteristika umfangreicher multivariater Datensätze erkennen, mittels geeigneter Methoden die Anzahl der zu analysierenden Dimensionen reduzieren sowie mittels Klassifikations- und Assoziationsverfahren Ähnlichkeiten und komplexe Zusammenhänge in den Daten offenlegen. Sie sind in der Lage, die Statistiksoftwarepakete R und SPSS als Werkzeuge statistischer Analysen einzusetzen und ihre erzielten Ergebnisse kritisch auszuwerten.</p>
<b>Inhalt</b>	<ul style="list-style-type: none"> <li>- Aufgaben, Ziele und Konzepte multivariater Analysen</li> <li>- Methoden der multivariaten Statistik: Hauptkomponenten- und Faktorenanalyse, Clusteranalyse, Diskriminanzanalyse</li> <li>- Methoden des Maschinenlernens: Klassifikationsbäume, Assoziationsregeln, Künstliche Neuronale Netze</li> </ul>
<b>Teilnahmevoraussetzungen</b>	keine
<b>Literaturangabe</b>	Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.
<b>Vergabe von Leistungspunkten</b>	Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Klausur 90 Min., mit Wichtung: 1</b>	
	E-Learning-Veranstaltung "Multivariate Datenanalyse und Data Mining" (4SWS)
	Übung "Multivariate Datenanalyse und Data Mining" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2104	Wahlpflicht

<b>Modultitel</b>	<b>Neuromorphe Informationsverarbeitung</b> Vertiefungsmodul
<b>Modultitel (englisch)</b>	Neuromorphic Information Processing In-Depth Module
<b>Empfohlen für:</b>	2. Semester
<b>Verantwortlich</b>	Professur für Neuromorphe Informationsverarbeitung
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	jedes Sommersemester
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• Vorlesung "Neuronal Computing" (2 SWS) = 30 h Präsenzzeit und 60 h Selbststudium = 90 h</li> <li>• Vorlesung "Neurobionische Systeme" (2 SWS) = 30 h Präsenzzeit und 60 h Selbststudium = 90 h</li> <li>• Praktikum "SNN" (2 SWS) = 30 h Präsenzzeit und 90 h Selbststudium = 120 h</li> </ul>
<b>Arbeitsaufwand</b>	10 LP = 300 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	<ul style="list-style-type: none"> <li>• M.Sc. Bioinformatik</li> <li>• M.Sc. Data Science</li> <li>• Vertiefungsmodul im M.Sc. Informatik der Technischen Informatik</li> <li>• Lehramt Informatik</li> <li>• M.Sc. Medizininformatik</li> <li>• Wahlpflichtmodul im M.Sc. Wirtschaftsinformatik</li> <li>• M.Sc. Wirtschaftspädagogik (zweites Fach Informatik)</li> </ul>
<b>Ziele</b>	<p>Nach der aktiven Teilnahme am Modul sind die Studierenden in der Lage</p> <ul style="list-style-type: none"> <li>- grundlegende Begriffe aus den beiden Vorlesungen zu definieren und zu erklären</li> <li>- ausgewählte Verfahren und Algorithmen zu beschreiben und zu analysieren</li> <li>- algorithmische Lösungsansätze zu erklären und diese selbstständig auf Problemstellungen anzuwenden</li> <li>- Aufgabenstellungen praktisch in Form eines Software-basierten Verfahrens zu lösen.</li> </ul>
<b>Inhalt</b>	<p>Vorlesung "Neuronal Computing"</p> <ul style="list-style-type: none"> <li>- Informationstheorie</li> <li>- Neurone als Rechner</li> <li>- Bidirektionale Kontaktierung von Neuronen</li> <li>- Signalverarbeitung von Nervensignalen</li> <li>- Modular und Population Coding</li> <li>- Unitary Events Analysis</li> <li>- Nerven-Maschine-Schnittstellen</li> </ul> <p>Vorlesung "Neurobionische Systeme"</p> <ul style="list-style-type: none"> <li>- Funktionsweise Neurone</li> <li>- Grundorganisation Gehirn</li> <li>- Funktionsweise Synapsen</li> <li>- Neuronale Netze</li> <li>- Selbstorganisation</li> </ul>

- Bioanaloge/Bioinspirierte neuronale Netze
- Anwendungen bionischer Systeme

Praktikum "SNN"

- Anwendung der wesentlichen Lösungsansätze auf Problemstellung der industriellen und wissenschaftlichen Anwendungen.

Als Praktikumsleistung stehen auf vielfachen Wunsch der Studierenden eine Projektarbeit in Gruppenarbeit bzw. 5 Versuche in Einzelarbeit zur Wahl. Die individuelle Festlegung hierzu erfolgt vor Beginn des Praktikums oder zu Beginn der Veranstaltung.

**Teilnahmevoraussetzungen**

keine

**Literaturangabe**

Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Mündliche Prüfung 25 Min., mit Wichtung: 1</b>	
<i>Prüfungsvorleistung: Praktikumsleistung (Bearbeitungsdauer 10 Wochen) mit Abschlusspräsentation (15 Minuten)</i>	
	Vorlesung "Neuronal Computing" (2SWS)
	Vorlesung "Neurobionische Systeme" (2SWS)
	Praktikum "SNN" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2218S	Wahlpflicht

<b>Modultitel</b>	<b>Grundlagen Komplexer Systeme (S)</b>
<b>Modultitel (englisch)</b>	Foundations of Complex Systems (S)
<b>Empfohlen für:</b>	2. Semester
<b>Verantwortlich</b>	Professur für Schwarmintelligenz und Komplexe Systeme
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	jedes Sommersemester
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• Vorlesung "Grundlagen Komplexer Systeme" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h</li> <li>• Seminar "Grundlagen Komplexer Systeme" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h</li> </ul>
<b>Arbeitsaufwand</b>	5 LP = 150 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	<ul style="list-style-type: none"> <li>• Kernmodul der Praktischen Informatik im M.Sc. Informatik</li> <li>• M.Sc. Data Science</li> <li>• Lehramt Informatik</li> </ul>
<b>Ziele</b>	<p>Nach der aktiven Teilnahme am Modul "Grundlagen Komplexer Systeme" sind die Studierenden in Lage:</p> <ul style="list-style-type: none"> <li>- grundlegende Begriffe beispielhafter komplexe Systeme zu formulieren und zu erklären,</li> <li>- für mindestens ein ausgewähltes komplexe System (z.B. Schwarmssysteme, diskrete Simulationssysteme oder Zellularautomaten) grundlegende Funktionsprinzipien zu analysieren und</li> <li>- diese Funktionsprinzipien selbstständig so einzusetzen, dass das System zur Lösung von Problemstellungen eingesetzt werden kann.</li> </ul>
<b>Inhalt</b>	<p>Es muss eine der folgenden Vorlesungen und das Seminar gewählt werden.</p> <p>Vorlesung "Diskrete Simulation": Simulationsparadigmen, Grundlagen von Warteschlangen/Bediensystemen, Formale Modelle für Diskrete Ereignissysteme und Systemspezifikation, Ein- und Ausgabemodellierung, Simulationssprachen, Parallele/Verteilte Simulation.</p> <p>Vorlesung "Zellularautomaten": Berechnungsmächtigkeit, Selbstreproduktion, Schnelles Sortieren, Synchronisations- und Markierungsprobleme, Diskretisierung kontinuierlicher Systeme, Modellierung realer Phänomene.</p> <p>Vorlesung "Verfahren der Schwarm Intelligenz": Ameisenalgorithmen, Schwarmalgorithmen, Prinzipien der Selbstorganisation in biologischen Systemen und ihre Nutzung in der Informatik.</p> <p>Seminar "Grundlagen Komplexer Systeme": Das Seminar behandelt ergänzende Arbeiten zu den Themen der Vorlesungen und befasst sich mit aktuellen wissenschaftlichen Abhandlungen aus den komplexen Systemen.</p>

**Teilnahmevoraussetzungen** Die Module "Grundlagen Komplexer Systeme (S)" (10-202-2218S) und "Grundlagen Komplexer Systeme (V)" (10-202-2218V) schließen sich gegenseitig aus.

**Literaturangabe** Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten** Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Mündliche Prüfung 20 Min., mit Wichtung: 1</b>	
	Vorlesung "Grundlagen Komplexer Systeme" (2SWS)
Referat (45 Min.) mit schriftlicher Ausarbeitung (2 Wochen), mit Wichtung: 1	Seminar "Grundlagen Komplexer Systeme" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2218V	Wahlpflicht

### Modultitel **Grundlagen Komplexer Systeme (V)**

**Modultitel (englisch)** Foundations of Complex Systems (V)

**Empfohlen für:** 2. Semester

**Verantwortlich** Professur für Schwarmintelligenz und Komplexe Systeme

**Dauer** 1 Semester

**Modulturnus** jedes Sommersemester

**Lehrformen**

- Vorlesung "Grundlagen Komplexer Systeme" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h
- Vorlesung mit seminaristischem Anteil "Grundlagen Komplexer Systeme 2" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h

**Arbeitsaufwand** 5 LP = 150 Arbeitsstunden (Workload)

**Verwendbarkeit**

- Kernmodul der Praktischen Informatik im M.Sc. Informatik
- M.Sc. Data Science
- Lehramt Informatik

**Ziele**

Nach der aktiven Teilnahme am Modul "Grundlagen Komplexer Systeme" sind die Studierenden in Lage:

- grundlegende Begriffe beispielhafter komplexe Systeme zu formulieren und zu erklären,
- für mindestens ein ausgewähltes komplexe System (z.B. Schwarmssysteme, diskrete Simulationssysteme oder Zellularautomaten) grundlegende Funktionsprinzipien zu analysieren und
- diese Funktionsprinzipien selbstständig so einzusetzen, dass das System zur Lösung von Problemstellungen eingesetzt werden kann.

**Inhalt**

Es muss die Vorlesung mit seminaristischem Anteil "Verfahren der Schwarmintelligenz" sowie eine der beiden Vorlesungen "Diskrete Simulation" oder "Zellularautomaten" gewählt werden.

Diskrete Simulation: Simulationsparadigmen, Grundlagen von Warteschlangen/Bediensystemen, Formale Modelle für Diskrete Ereignissysteme und Systemspezifikation, Ein- und Ausgabemodellierung, Simulationssprachen, Parallele/Verteilte Simulation.

Zellularautomaten: Berechnungsmächtigkeit, Selbstreproduktion, Schnelles Sortieren, Synchronisations- und Markierungsprobleme, Diskretisierung kontinuierlicher Systeme, Modellierung realer Phänomene.

Verfahren der Schwarm Intelligenz: Ameisenalgorithmen, Schwarmalgorithmen, Prinzipien der Selbstorganisation in biologischen Systemen und ihre Nutzung in der Informatik

**Teilnahmevoraussetzungen**

Die Module "Grundlagen Komplexer Systeme (S)" (10-202-2218S) und "Grundlagen Komplexer Systeme (V)" (10-202-2218V) schließen sich gegenseitig aus.

**Literaturangabe** Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten** Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Mündliche Prüfung 20 Min., mit Wichtung: 1</b>	
	Vorlesung "Grundlagen Komplexer Systeme" (2SWS)
	Vorlesung mit seminaristischem Anteil "Grundlagen Komplexer Systeme 2" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2346	Wahlpflicht

<b>Modultitel</b>	<b>Automated Software Engineering</b> Vertiefungsmodul
<b>Modultitel (englisch)</b>	Automated Software Engineering In-Depth Module
<b>Empfohlen für:</b>	2. Semester
<b>Verantwortlich</b>	Professur für Softwaresysteme
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	jedes Sommersemester
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• E-Learning-Veranstaltung "Automated Software Engineering" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h</li> <li>• Übung "Automated Software Engineering" (1 SWS) = 15 h Präsenzzeit und 60 h Selbststudium = 75 h</li> <li>• Seminar "Advanced Software Engineering" (2 SWS) = 30 h Präsenzzeit und 120 h Selbststudium = 150 h</li> </ul>
<b>Arbeitsaufwand</b>	10 LP = 300 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	<ul style="list-style-type: none"> <li>• Vertiefungsmodul M.Sc. Informatik</li> <li>• Wahlpflichtmodul M.Sc. Data Science</li> <li>• Ergänzungsbereich</li> </ul>
<b>Ziele</b>	<p>Nach der aktiven Teilnahme am Modul Automated Software Engineering sind die Studierenden in der Lage verschiedene Aktivitäten in der Softwareentwicklung zu optimieren und automatisieren. So sind Studierende in der Lage Tests automatisiert zu generieren, optimale Softwarekonfigurationen automatisiert durch den Einsatz von Metaheuristiken zu finden und kleine Funktionen und Code selbst zu synthetisieren.</p> <p>Durch eine Mischung an Lehrformen werden unterschiedliche Qualifikationen erreicht. Die Videovorlesung "Automated Software Engineering" befähigt die Studierende Optimierungs- und Suchprobleme im Feld des Software Engineerings zu erkennen, auszuformulieren und zu lösen. Sie sind weiterhin in der Lage selbst Algorithmen zu implementieren und Zielfunktionen zu definieren. Darüber hinaus werden sie befähigt Analysen von Softwaresystemen bzw. Repositories durchzuführen für einen möglichen KI-Einsatz im Software Engineering.</p>
<b>Inhalt</b>	<p>Das Modul stellt Konzepte, Theorien und Algorithmen für metaheuristische Optimierungsverfahren vor:</p> <ul style="list-style-type: none"> <li>- Single-State Optimization (z.B., Hill Climbing, Simulated Annealing)</li> <li>- Multi-State Optimization (z.B., evolutionäre Algorithmen)</li> <li>- Multi-Dimensional Optimization (z.B. NSGA-II, IBEA)</li> <li>- Constraint Satisfaction Problem Solving</li> </ul> <p>Verzahnt werden diese Inhalte mit Themen des Software Engineerings (Konfigurationsoptimierung, Testgenerierung, Codesynthese etc.) bei denen Suchbasierte Lösungen erforderlich sind. Hierdurch werden Kompetenzen zur Automatisierung im Software Engineering gebildet.</p>

Darüber hinaus bzw. alternativ in Ausrichtung an neuesten wissenschaftlichen Erkenntnissen werden Grundlagen im Bereich von Software- und/oder Repositoryanalysen vermitteln.

Das Seminar stellt mehrere, wechselnde Themenkomplexe bei der Automatisierung des Software Engineerings dar, wie z.B. Automatic Code Repair, Code Summarization and Completion, Automatically Finding Bugs, Finding Optimal Software Configurations, Replikation von Forschungsvorhaben in diesem Gebiet, usw. Basierend auf diesem Thema erarbeiten die Studierenden den Stand der Forschung und implementieren und reproduzieren bzw. rekombinieren und entwickeln neuartige Ansätze.

**Teilnahmevoraussetzungen**

Teilnahme am Modul "Künstliche Neuronale Netze, Deep Learning, Maschinelles Lernen und Signalverarbeitung" oder "Grundlagen des maschinellen Lernens" nicht kombinierbar mit Modul "Grundlagen des Automated Software Engineering"

**Literaturangabe**

Handbook of Metaheuristics, Fred W. Glover, Gary A. Kochenberger, Springer Science & Business Media, 2006.

Machine Learning, Tom Mitchell, McGraw-Hill Education, 1997.

Essentials of Metaheuristics, Sean Luke, 2013.

A Field Guide to Genetic Programming, Riccardo Poli, William B. Langdon, Nicholas Freitag McPhee, Lulu Pr, 2008.

Weitere Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Klausur 90 Min., mit Wichtung: 1</b>	
<i>Prüfungsvorleistung: Referat mit schriftlicher Ausarbeitung (30 Minuten), Bearbeitungszeit: 8 Wochen</i>	
	E-Learning-Veranstaltung "Automated Software Engineering" (2SWS)
	Übung "Automated Software Engineering" (1SWS)
	Seminar "Advanced Software Engineering" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-202-2350	Wahlpflicht

<b>Modultitel</b>	<b>Grundlagen des Automated Software Engineerings</b> Kernmodul
<b>Modultitel (englisch)</b>	Fundamentals of Automated Software Engineering Key Module
<b>Empfohlen für:</b>	2. Semester
<b>Verantwortlich</b>	Professur für Softwaresysteme
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	jedes Sommersemester
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• Vorlesung "Automated Software Engineering" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h</li> <li>• Übung "Automated Software Engineering" (1 SWS) = 15 h Präsenzzeit und 60 h Selbststudium = 75 h</li> </ul>
<b>Arbeitsaufwand</b>	5 LP = 150 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	<ul style="list-style-type: none"> <li>• Kernmodul M.Sc. Informatik im Bereich Praktische Informatik</li> <li>• Wahlpflichtmodul M.Sc. Data Science, Ergänzungsbereich</li> </ul>
<b>Ziele</b>	<p>Nach der aktiven Teilnahme am Modul Automated Software Engineering sind die Studierenden in der Lage verschiedene Aktivitäten in der Softwareentwicklung zu optimieren und automatisieren. So sind Studierende in der Lage Tests automatisiert zu generieren, optimale Softwarekonfigurationen automatisiert durch den Einsatz von Metaheuristiken zu finden und kleine Funktionen und Code selbst zu synthetisieren.</p> <p>Durch eine Mischung an Lehrformen werden unterschiedliche Qualifikationen erreicht. Die Videovorlesung "Automated Software Engineering" befähigt die Studierende Optimierungs- und Suchprobleme im Feld des Software Engineerings zu erkennen, auszuformulieren und zu lösen. Sie sind weiterhin in der Lage selbst Algorithmen zu implementieren und Zielfunktionen zu definieren. Darüber hinaus werden sie befähigt Analysen von Softwaresystemen bzw. Repositories durchzuführen für einen möglichen KI-Einsatz im Software Engineering.</p>
<b>Inhalt</b>	<p>Das Modul stellt Konzepte, Theorien und Algorithmen für metaheuristische Optimierungsverfahren vor:</p> <ul style="list-style-type: none"> <li>- Single-State Optimization (z.B., Hill Climbing, Simulated Annealing)</li> <li>- Multi-State Optimization (z.B., evolutionäre Algorithmen)</li> <li>- Multi-Dimensional Optimization (z.B. NSGA-II, IBEA)</li> <li>- Constraint Satisfaction Problem Solving</li> </ul> <p>Verzahnt werden diese Inhalte mit Themen des Software Engineerings (Konfigurationsoptimierung, Testgenerierung, Codesynthese etc.) bei denen Suchbasierte Lösungen erforderlich sind. Hierdurch werden Kompetenzen zur Automatisierung im Software Engineering gebildet.</p> <p>Darüber hinaus bzw. alternativ in Ausrichtung an neuesten wissenschaftlichen Erkenntnissen werden Grundlagen im Bereich von Software- und/oder Repositoryanalysen vermittelt.</p>

**Teilnahmevoraussetzungen** Gegenseitiger Ausschluss des Kernmoduls "Grundlagen des Automated Software Engineering" und des Vertiefungsmoduls "Automated Software Engineering" (10-202-2346).

**Literaturangabe** Handbook of Metaheuristics, Fred W. Glover, Gary A. Kochenberger, Springer Science & Business Media, 2006.  
 Essentials of Metaheuristics, Sean Luke, 2013.  
 A Field Guide to Genetic Programming, Riccardo Poli, William B. Langdon, Nicholas Freitag McPhee, Lulu Pr, 2008.

Weitere Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten** Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Klausur 90 Min., mit Wichtung: 1</b>	
	Vorlesung "Automated Software Engineering" (2SWS)
	Übung "Automated Software Engineering" (1SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-DIH-1001	Wahlpflicht

### Modultitel **Introduction to Linguistic Annotation and XML Technologies**

**Modultitel (englisch)** Introduction to Linguistic Annotation and XML Technologies

**Empfohlen für:** 2. Semester

**Verantwortlich** Leitung des Instituts für Informatik

**Dauer** 1 Semester

**Modulturnus** unregelmäßig

**Lehrformen**

- Vorlesung "Einführung in linguistische Annotation und XML Technologien" (2 SWS) = 30 h Präsenzzeit und 120 h Selbststudium = 150 h
- Praktikum "Praktikum" (0 SWS) = 0 h Präsenzzeit und 150 h Selbststudium = 150 h

**Arbeitsaufwand** 10 LP = 300 Arbeitsstunden (Workload)

**Verwendbarkeit**

- M.Sc. Data Science
- M.Sc. Digital Humanities
- M.Sc. Informatik

**Ziele**

Nach der aktiven Teilnahme am Modul "Einführung in linguistische Annotation und XML Technologien" sind die Studierende in der Lage, die Hauptkomponenten der linguistischen Annotation zu erkennen und zu definieren. Dazu gehören die Kenntnis der Konzepte Notationsschema, Annotation, Training- und Testphasen und Auswertung, sowie die Fähigkeit, den Unterschieden zwischen die wichtigsten XML Technologien (XML Schema, XQuery, XSLT) zu erklären. Die Studierenden können einfache Texte gemäß den TEI-Richtlinien interpretieren und kodieren und Unicode-bezogene Probleme lösen. Sie können einfache XQuery/XSLT Skripte schreiben, um Texte abzufragen und zu transformieren.

**Inhalt**

Vorlesung:

- 1 Prinzipien der manuellen Annotation und Maschinenlernen
- 2 Text encoding: Dateien-Formate (TXT, XML, JSON), UNICODE, TEI (= Text Encoding Initiative)
- 3 XML und XML Schema foundations
- 4 XQuery: Extraktion und Manipulation von Informationen
- 5 XSLT: Transformation von XML Dokumenten

Praktikum:

- 1 Übungen zum Kodieren von Text, Annotation, POS Training, Taggers/Syntaktische Parsers sowie Query/Transformation von Dokumenten

Die Lehrveranstaltungen werden in englischer Sprache gehalten. Studien- und Prüfungsleistungen sind in englischer Sprache zu erbringen.

**Teilnahmevoraussetzungen** keine

**Literaturangabe** Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben.  
Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Elektronische Prüfung 90 Min., mit Wichtung: 1</b>	
	Vorlesung "Einführung in linguistische Annotation und XML Technologien" (2SWS)
	Praktikum "Praktikum" (0SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-DIH-1002	Wahlpflicht

### Modultitel **Computational Methods for Empirical Research**

**Modultitel (englisch)** Computational Methods for Empirical Research

**Empfohlen für:** 2./3. Semester

**Verantwortlich** Professur für Computational Humanities

**Dauer** 1 Semester

**Modulturnus** mindestens einmal alle 2 Jahre

**Lehrformen**

- Vorlesung "Computational Methods for Empirical Research" (2 SWS) = 30 h Präsenzzeit und 90 h Selbststudium = 120 h
- Seminar "Computational Methods for Empirical Research" (2 SWS) = 30 h Präsenzzeit und 150 h Selbststudium = 180 h

**Arbeitsaufwand** 10 LP = 300 Arbeitsstunden (Workload)

**Verwendbarkeit**

- M.Sc. Data Science: Ergänzungsbereich
- M.Sc. Digital Humanities
- M.Sc. Informatik: Ergänzungsbereich

**Ziele**

Nach der aktiven Teilnahme am Modul "Computational Methods for Empirical Research" beherrschen die Studierenden folgende Kompetenzen:

- Anwendung der Grundlagen empirischer, inhaltsbasierter Forschung
- Nutzung computergestützter Verfahren für empirische Forschungsabläufe
- Operationalisierung empirischer Forschungsfragen mit computerbasierten Verfahren
- Reflexion theoretischer Grundannahmen, Forschungsdesigns und -erkenntnisse
- Grundverständnis computergestützter Mixed Methods

**Inhalt**

Das Modul besteht aus einer Vorlesung und einem begleitenden Seminar. In der Vorlesung werden die theoretischen Grundlagen für computergestützte Tools und Methoden in der empirischen Forschung vermittelt. Im Seminar werden diese Grundlagen dann anhand von Lektüretexten, praktischen Aufgaben und Referaten weiter vertieft. Die erlernten Fertigkeiten werden in einer eigenständigen Projektarbeit synthetisiert.

In diesem Modul erlernen die Studierenden computerbasierte Tools und Methoden, um aus diversen Inhaltstypen automatisch Ausprägungen von Variablen zu ermitteln. Dabei werden konkret Methoden der Mustererkennung, des maschinellen Lernens und der künstlichen Intelligenz vorgestellt und gelehrt, wie empirische Fragestellungen mit diesen Methoden operationalisiert werden. Hier wird besonderer Wert auf Multimodalität (Text, Bild, Ton) und die Einbettung theoretischer Vorannahmen Wert gelegt. Die Inhalte lassen sich folgendermaßen beschreiben:

- Die Rolle von Theorie in der empirischen inhaltsbasierten Forschung
- Forschungsdesign und typische Workflows
- Mixed Methods (Kombination aus qualitativen und quantitativen Methoden)
- Messvariablen, Klassen, Kategorien
- Messfehler
- Annotation
- Hypothesen

- Operationalisierung
- Medientypen, Inhalte
- Computerbasierte Verfahren zur multimodalen und automatischen Extraktion von Variablen
- Mustererkennung, Maschinelles Lernen, künstliche Intelligenz
- Beurteilung von Datenqualität

**Teilnahmevoraussetzungen**      Algorithmen und Datenstrukturen 1+2 oder 10 Leistungspunkte in der angewandten Informatik

**Literaturangabe**              Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**      Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Projektarbeit: Projektbericht (8 Wochen) und Präsentation (30 Min.), mit Wichtung: 1</b>	
	Vorlesung "Computational Methods for Empirical Research" (2SWS)
	Seminar "Computational Methods for Empirical Research" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-DIH-1003	Wahlpflicht

### Modultitel **Computational Spatial Humanities**

**Modultitel (englisch)** Computational Spatial Humanities

**Empfohlen für:** 2./3. Semester

**Verantwortlich** Professur für Computational Humanities

**Dauer** 1 Semester

**Modulturnus** mindestens einmal alle 2 Jahre

**Lehrformen**

- Vorlesung "Computational Spatial Humanities" (2 SWS) = 30 h Präsenzzeit und 90 h Selbststudium = 120 h
- Seminar "Computational Spatial Humanities" (2 SWS) = 30 h Präsenzzeit und 150 h Selbststudium = 180 h

**Arbeitsaufwand** 10 LP = 300 Arbeitsstunden (Workload)

**Verwendbarkeit**

- M.Sc. Data Science: Ergänzungsbereich
- M.Sc. Digital Humanities
- M.Sc. Informatik: Ergänzungsbereich

**Ziele**

Ziel dieses Moduls ist die Vermittlung von Kompetenzen in der computergestützten Auswertung raumbezogener Daten im Kontext geisteswissenschaftlicher Fragestellungen. Die erlernten Fähigkeiten und ein praxisnaher Einblick in digitale geographische Forschung befähigen zur späteren interdisziplinären wissenschaftlichen Arbeit. Diese kann sich etwa im Bereich der Globalisierungsforschung anschließen welche mit dem ReCentGlobe einen wichtigen Schwerpunkt am Forschungsstandort Leipzig darstellt.

Nach erfolgreicher Teilnahme am Modul sind die Studierenden in der Lage,

- grundlegende Begriffe der Spatial Humanities zu definieren und die dahinterliegenden Konzepte zu erklären,
- Methoden und Werkzeuge des raumbezogenen Distant Readings sowie der geographischen Informationssysteme (GIS) sicher und zweckmäßig anzuwenden,
- realweltliche Datensätze mit Raumbezug zu analysieren und sie im Kontext ihrer Entstehung zu interpretieren
- wissenschaftliche Ansätze der Spatial Humanities für die Eignung zur Beantwortung spezifischer Fragestellungen zu bewerten

**Inhalt**

- Abgrenzung physischer Geographie von Humangeographie und Überblick zu etablierten, digitale Methoden
- Einführung in Raumkonzepte, Raumbegriffe und Verräumlichungsprozesse
- Grundlagen der Kartographie (Koordinatensysteme, Kartenprojektionen, Zweck und Formsprache von Karten)
- Grundlagen Geographischer Informationssysteme (Datenformate, Transformationen, Abfragen, Datenintegration)
- Grundlagen der Geo-Visualisierung in den digitalen Geisteswissenschaften (Vorgehensweise, Werkzeuge, Abbildung von Granularität, Unschärfe, Datenprovenienz)
- Multimediale Quellen für raumbezogene Forschung

- Georeferenzierung, Geotagging und Geo-Annotation zur Erschließung von Quellenmaterial
- Statistische und computergestützte qualitative Auswertung geotemporaler Datensätze (Einbeziehung von Vorannahmen, visuelle Autokorrelationen und Rahmenbedingung für die Interpretation)
- Grundlagen von Deep Maps

Im Seminar werden die Vorlesungsinhalte durch Gruppenarbeit sowie Vorträge gemeinschaftlich um praxisnahe Themen ergänzt.  
Die erlernten Fertigkeiten werden in einer eigenständigen Projektarbeit synthetisiert.

**Teilnahmevoraussetzungen**

keine

**Literaturangabe**

Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben.  
Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Projektarbeit: Projektbericht (8 Wochen) und Präsentation (30 Min.), mit Wichtung: 1</b>	
	Vorlesung "Computational Spatial Humanities" (2SWS)
	Seminar "Computational Spatial Humanities" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-DIH-1004	Wahlpflicht

### Modultitel **Cultural Analytics**

**Modultitel (englisch)** Cultural Analytics

**Empfohlen für:** 2./3. Semester

**Verantwortlich** Professur für Computational Humanities

**Dauer** 1 Semester

**Modulturnus** mindestens einmal alle 2 Jahre

**Lehrformen**

- Vorlesung "Cultural Analytics" (2 SWS) = 30 h Präsenzzeit und 90 h Selbststudium = 120 h
- Seminar "Cultural Analytics" (2 SWS) = 30 h Präsenzzeit und 150 h Selbststudium = 180 h

**Arbeitsaufwand** 10 LP = 300 Arbeitsstunden (Workload)

**Verwendbarkeit**

- M.Sc. Data Science: Ergänzungsbereich
- M.Sc. Digital Humanities
- M.Sc. Informatik: Ergänzungsbereich

**Ziele**

Nach der aktiven Teilnahme am Modul "Cultural Analytics" beherrschen die Studierenden folgende Kompetenzen:

- Überblick zu Forschungsparadigmen der Cultural Analytics
- Reflexion theoretischer Grundlagen von Cultural Analytics
- Anwendung wissenschaftlicher Standards und Methoden
- Operationalisierung von Forschungsfragen und Konzeption von Forschungsdesigns
- Nutzung computergestützter Verfahren für die Analyse und Visualisierung von Kulturdaten

**Inhalt**

In zunehmendem Maße wurden in den letzten Jahren ganz unterschiedliche Kulturartefakte wie Bücher, Zeitschriften, Bilder, Games, Filme und Musik massendigitalisiert. Damit ergeben sich aus wissenschaftlicher Perspektive ganz neue Möglichkeiten der quantitativen Analyse und explorativen Visualisierung, die häufig als Cultural Analytics zusammengefasst werden.

Das Modul besteht aus einer Vorlesung und einem begleitenden Seminar. In der Vorlesung werden die Grundlagen der „Cultural Analytics“ vermittelt. Im Seminar werden diese Grundlagen dann anhand von Lektüretexten, praktischen Aufgaben und Referaten weiter vertieft. Die erlernten Fertigkeiten werden in einer eigenständigen Projektarbeit synthetisiert.

Die Inhalte lassen sich folgendermaßen beschreiben:

- Theoretische Grundlagen der Cultural Analytics
- Diskussion bestehender Cultural Analytics-Studien
- Planung und Durchführung von Experimenten
- Methoden des Humanities Data Mining
- Informationsvisualisierung und Visual Analytics
- Interpretation quantitativer Muster in kulturellen Korpora

**Teilnahmevoraussetzungen** keine

**Literaturangabe** Manovich, Lev (2020). Cultural Analytics. MIT Press.  
Weitere Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten** Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben.  
Näheres regelt die Prüfungsordnung.

### **Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Projektarbeit: Projektbericht (8 Wochen) und Präsentation (30 Min.), mit Wichtung: 1</b>	
	Vorlesung "Cultural Analytics" (2SWS)
	Seminar "Cultural Analytics" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-INF-DS101	Wahlpflicht

### Modultitel **Skalierbare Datenbanktechnologien 2**

**Modultitel (englisch)** Scalable Database Technologies 2

**Empfohlen für:** 2. Semester

**Verantwortlich** Professur für Informatik (Datenbanken)

**Dauer** 1 Semester

**Modulturnus** jedes Sommersemester

**Lehrformen**

- Vorlesung "Skalierbare Datenbanktechnologien 3" (2 SWS) = 30 h Präsenzzeit und 30 h Selbststudium = 60 h
- Vorlesung "NoSQL-Datenbanken" (1 SWS) = 15 h Präsenzzeit und 30 h Selbststudium = 45 h
- Übung "NoSQL-Datenbanken" (1 SWS) = 15 h Präsenzzeit und 30 h Selbststudium = 45 h

**Arbeitsaufwand** 5 LP = 150 Arbeitsstunden (Workload)

**Verwendbarkeit** • M.Sc. Data Science

**Ziele** Nach der aktiven Teilnahme am Modul "Skalierbare Datenbanktechnologien 2" sind die Studierenden in der Lage verschiedene Architekturen aktueller Datenbank Anwendungen anzugeben.

Sie können Verfahren zur Verarbeitung großer Datenmengen erläutern.

Die Studierenden sind imstande aktuelle Datenbanktechnologien und Verfahren zur Verarbeitung und Analyse großer Datenmengen anzuwenden.

Sie sind in der Lage, selbstständig Anfragen durchzuführen.

### Inhalt

Vorlesung NoSQL-Datenbanken

- Verwaltung großer Datenmengen in verteilten Clusterumgebungen
- Kategorisierung und Eigenschaften von NoSQL-Datenbanksystemen
- Vergleich von NoSQL-Systemen zu relationalen Datenbanksystemen
- Partitionierung, Konsistenz, Replikation
- Key-Value, Document Stores und Extensible Record Stores
- Graphdatenbanken

Übung NoSQL-Datenbanken

- Verständnisaufgaben zum Inhalt der Vorlesung
- Praktische Arbeit mit realen, teilweise sehr großen, Datensätzen
- Formulierung und Ausführung von Anfragen in den jeweiligen Sprachen verschiedener Datenbankimplementierungen
- Umgang mit den Anbindungen an gängige Programmiersprachen

Außerdem wird mindestens eine der folgenden Vorlesungen angeboten. Eine Vorlesung wird ausgewählt.

Vorlesung Data Warehousing

- Architektur von Data Warehouse-Systemen
- Mehrdimensionale Modellierung
- Datenintegration, Datenbereinigung, ETL-Werkzeuge
- Performance-Techniken: Indexstrukturen, materialisierte Sichten, parallele Datenbanken
- Data Mining-Verfahren
- Anwendungen von Datawarehouses

Vorlesung Implementierung von Datenbanksystemen II

- Synchronisation: Serialisierbarkeit, Sperrverfahren, Deadlock-Behandlung,
- Mehrversionenverfahren, sonstige Synchronisationsansätze
- Logging und Recovery: Fehlermodell, Logging-Strategien, Checkpoint-Ansätze, Crash-Recovery, Media-Recovery
- Erweiterte Transaktionsmodelle (geschachtelte Transaktionen, verkettete Transaktionen, etc.)
- DB-Benchmarks.

Vorlesung Datenintegration

- Überblick zur Integration verteilter, heterogener Datenbestände
- Verteilung, Autonomie und Heterogenität
- Eigenschaften von Integrationssystemen
- Architekturen von Integrationssystemen
- Anfrageverarbeitung
- Schemamanagement
- Datenfusion

**Teilnahmevoraussetzungen** keine

**Literaturangabe** Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten** Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Klausur 120 Min., mit Wichtung: 1</b>	
<i>Prüfungsvorleistung: Testat (60 Min) in der Übung "NoSQL-Datenbanken"</i>	
	Vorlesung "Skalierbare Datenbanktechnologien 3" (2SWS)
	Vorlesung "NoSQL-Datenbanken" (1SWS)
	Übung "NoSQL-Datenbanken" (1SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-INF-DS102	Wahlpflicht

### Modultitel **Big Data Praktikum**

**Modultitel (englisch)** Big Data Project

**Empfohlen für:** 2. Semester

**Verantwortlich** Professur für Informatik (Datenbanken)

**Dauer** 1 Semester

**Modulturnus** jedes Sommersemester

**Lehrformen** • Praktikum "Big Data Praktikum" (2 SWS) = 30 h Präsenzzeit und 120 h Selbststudium = 150 h

**Arbeitsaufwand** 5 LP = 150 Arbeitsstunden (Workload)

**Verwendbarkeit** • M.Sc. Data Science

**Ziele** Nach der aktiven Teilnahme am Modul "Big Data Praktikum" sind die Studierenden in der Lage, aktuelle Datenbank-/Big-Data-Technologien anzuwenden.

Sie können selbständig Anfragen formulieren und ausführen.

Die Studierenden sind imstande, Algorithmen zur effizienten Verarbeitung großer Datenmengen zu implementieren.

**Inhalt**

- Entwurf und Realisierung einer Anwendung oder eines Algorithmus unter Verwendung von existierenden Big Data Frameworks und Technologien,
- Anfertigung eines Entwurfs zur Darstellung des Konzepts und der Architektur zur geplanten Implementierung unter Einbeziehung des jeweils verwendeten Frameworks,
- Implementierung einer Anwendung und zur Verfügung stellen eines dokumentierten, ausführbaren Programms oder Ablaufplans,
- Präsentation der Anwendung und Ergebnisse des Praktikums

**Teilnahmevoraussetzungen** keine

**Literaturangabe** Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten** Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

### Prüfungsleistungen und -vorleistungen

<b>Modulprüfung: Praktikumsleistung (3 Testate a 60 Min.), mit Wichtung: 1</b>	
	Praktikum "Big Data Praktikum" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-INF-DS104	Wahlpflicht

<b>Modultitel</b>	<b>Data Preparation &amp; Cleaning</b>
<b>Modultitel (englisch)</b>	Data Preparation & Cleaning
<b>Empfohlen für:</b>	2./4. Semester
<b>Verantwortlich</b>	Professur für Informatik (Datenbanken)
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	jedes Sommersemester
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• Vorlesung "Data Wrangling" (2 SWS) = 30 h Präsenzzeit und 60 h Selbststudium = 90 h</li> <li>• Praktikum "Data Wrangling" (1 SWS) = 15 h Präsenzzeit und 45 h Selbststudium = 60 h</li> </ul>
<b>Arbeitsaufwand</b>	5 LP = 150 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	• M.Sc. Data Science: Bereich Skalierbares Datenmanagement
<b>Ziele</b>	<p>Nach der Teilnahme an diesem Modul sind die Studenten in der Lage die notwendigen Prozesse der Datenaufbereitung zu erläutern. Hierfür können sie prozessspezifische Verfahren bzgl. der Datenextraktion, Qualitätsbewertung, und Datenbereinigung nennen und erklären. Weiterhin sind die Studierenden in der Lage, Verfahren der Datenintegration anzuwenden und zu realisieren.</p>
<b>Inhalt</b>	<p>Dieses Modul befasst sich mit der Analyse und Aufbereitung von Daten, um diese für anwendungsspezifische Analysen bereitzustellen. Aufgrund existierender Datenqualitätsprobleme ist eine Qualitätsanalyse und die sich daraus ergebene Datenbereinigung sowie Datenintegration unabdingbar. Hierfür finden diverse Verfahren Anwendung u.a. Machine-Learning basierte Verfahren für die Datenintegration.</p> <p>- Lehrsprache: englisch oder deutsch  - Prüfungssprache: englisch oder deutsch  Die Festlegung hierzu erfolgt vor der Moduleinschreibung auf elektronischem Weg (Vorlesungsverzeichnis) oder zu Beginn der Veranstaltung durch den Dozenten/die Dozentin.</p>
<b>Teilnahmevoraussetzungen</b>	keine
<b>Literaturangabe</b>	Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.
<b>Vergabe von Leistungspunkten</b>	Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Klausur 60 Min., mit Wichtung: 1</b>	
	Vorlesung "Data Wrangling" (2SWS)
Praktikumsleistung (4 Testate (Bearbeitungsdauer je 3 Wochen)), mit Wichtung: 1	Praktikum "Data Wrangling" (1SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-INF-DS105	Wahlpflicht

<b>Modultitel</b>	<b>Datenschutz kompakt von Anonymisierung bis Zweckbindung</b>
<b>Modultitel (englisch)</b>	Compact Data Privacy from Anonymization to Purpose Limitation
<b>Empfohlen für:</b>	2./4. Semester
<b>Verantwortlich</b>	Professur für Data Privacy and Security
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	jedes Sommersemester
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• Vorlesung "Datenschutz von Anonymisierung bis Zweckbindung" (2 SWS) = 30 h Präsenzzeit und 60 h Selbststudium = 90 h</li> <li>• Übung "Anwendung von Datenschutzverfahren" (1 SWS) = 15 h Präsenzzeit und 45 h Selbststudium = 60 h</li> </ul>
<b>Arbeitsaufwand</b>	5 LP = 150 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	<ul style="list-style-type: none"> <li>• M.Sc. Data Science: Bereich Skalierbares Datenmanagement</li> <li>• M.Sc. Informatik: Kernmodul der Praktischen Informatik oder Angewandten Informatik</li> </ul>
<b>Ziele</b>	<p>Nach der aktiven Teilnahme am Modul "Datenschutz kompakt von Anonymisierung bis Zweckbindung" verfügen die Studierenden über ein umfassendes Verständnis der gesellschaftlichen, technischen und wirtschaftlichen Ziele des Datenschutzes. Sie erwerben detaillierte theoretische und praktische Kenntnisse von grundlegenden Anonymisierungsverfahren und Anonymitätsmaßen. Dazu gehört auch die Abgrenzung der wesentlichen Fachbegriffe des Datenschutzes, insbesondere von denen aus der Datensicherheit. Die Studierenden können Datenschutzlösungen auf unterschiedlichen technischen und organisatorischen Ebenen anwenden und bewerten, sowie Systeme nach dem Standard-Datenschutzmodell absichern. Sie können technische Datenschutzlösungen im Internet zielführend einsetzen, z.B. Mix-Kaskaden, TOR, Freenet oder DoNotTrack.</p>
<b>Inhalt</b>	<p>Die Vorlesung führt in die gesellschaftlichen Grundlagen und Zielstellungen des Datenschutzes ein. Sie bietet einen breit angelegten Überblick über technische Lösungen des Datenschutzes mit Fokus auf Datenschutz im Internet. Zu diesen Lösungen zählen Anonymisierungsverfahren, Verfahren zum Verbergen der IP-Adresse oder der Urheberschaft von Informationen, sowie Ansätze zum Datenselbstschutz. In der Übung werden diese Konzepte an praktischen Beispielen vertieft.</p>
<b>Teilnahmevoraussetzungen</b>	<p>Teilnahme am Modul "Betriebs- und Kommunikationssysteme" (10-201-2004) oder gleichwertige Kenntnisse. Die Module 10-INF-DS105 und 10-INF-DS107 schließen sich gegenseitig aus.</p>
<b>Literaturangabe</b>	Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.
<b>Vergabe von Leistungspunkten</b>	Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Klausur 60 Min., mit Wichtung: 1</b>	
	Vorlesung "Datenschutz von Anonymisierung bis Zweckbindung" (2SWS)
	Übung "Anwendung von Datenschutzverfahren" (1SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-INF-DS106	Wahlpflicht

<b>Modultitel</b>	<b>Verschlüsseltes Datenmanagement auf privaten Daten</b>
<b>Modultitel (englisch)</b>	Encrypted Data Management on Private Data
<b>Empfohlen für:</b>	2./4. Semester
<b>Verantwortlich</b>	Professur für Data Privacy and Security
<b>Dauer</b>	1 Semester
<b>Modulturnus</b>	jedes Sommersemester
<b>Lehrformen</b>	<ul style="list-style-type: none"> <li>• Vorlesung "Verschlüsseltes Datenmanagement auf privaten Daten" (2 SWS) = 30 h Präsenzzeit und 60 h Selbststudium = 90 h</li> <li>• Seminar "Neue Verfahren für verschlüsseltes Datenmanagement" (1 SWS) = 15 h Präsenzzeit und 45 h Selbststudium = 60 h</li> </ul>
<b>Arbeitsaufwand</b>	5 LP = 150 Arbeitsstunden (Workload)
<b>Verwendbarkeit</b>	• M.Sc. Data Science: Bereich Skalierbares Datenmanagement
<b>Ziele</b>	Nach der aktiven Teilnahme am Modul „Verschlüsseltes Datenmanagement auf privaten Daten“ verfügen die Studierenden über die erforderlichen theoretische und praxisbezogene Fähigkeiten zur Analyse, Implementierung und Bewertung von technischen Lösungen zur sicheren, datenschutzkonformen Auslagerung von Datenbankinhalten an Dritte bzw. in die Cloud. Dazu gehören vertiefte Kenntnisse der Sicherheitsgarantien, die technische Lösungen wie verschlüsselte Indexte, verschlüsselte Bucketization, Fragmentierung, Blockchain etc. bieten können. Die Studierenden können eine sinnvolle Zuordnung von Angreifermodellen zu Einsatzszenarien und dazu passenden Schutzmaßnahmen vornehmen. Sie besitzen die Kenntnisse zur Integration von Schutzmaßnahmen in bestehende Systeme unter Berücksichtigung von Sicherheits- und Performanzkriterien, sowie die Fähigkeit zur strukturierten Präsentation derartiger Aspekte.
<b>Inhalt</b>	Im Fokus dieses Moduls steht das Auslagern von Datenbankinhalten an Dritte, z.B. im Rahmen von Cloud-Lösungen, Containervirtualisierung oder Database-as-a-Service. In der Vorlesung werden dazu zunächst spezifische Angreifermodelle und Einsatzszenarien thematisiert. Diesen werden dann passende technische Maßnahmen wie verschlüsselte Indexte oder Blockchain gegenübergestellt. Im Seminarteil des Moduls müssen die Studierenden diese Kenntnisse anwenden, um neuartige Schutzmaßnahmen in diesem Fachgebiet zu präsentieren und kritisch zu hinterfragen, bzw. um neuartige Einsatzszenarien auf ihre Absicherbarkeit mit bekannten Maßnahmen hin zu bewerten.
<b>Teilnahmevoraussetzungen</b>	Teilnahme am Modul "Skalierbare Datenbanktechnologien 1" (10-INF-DS01) oder gleichwertige Kenntnisse
<b>Literaturangabe</b>	Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.
<b>Vergabe von Leistungspunkten</b>	Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Klausur 60 Min., mit Wichtung: 1</b>	
<i>Prüfungsvorleistung: Referat (15 Min.) im Seminar</i>	
	Vorlesung "Verschlüsseltes Datenmanagement auf privaten Daten" (2SWS)
	Seminar "Neue Verfahren für verschlüsseltes Datenmanagement" (1SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-INF-DS107	Wahlpflicht

### Modultitel **Datenschutz von Anonymisierung bis Zweckbindung**

**Modultitel (englisch)** Data Privacy from Anonymization to Purpose Limitation

**Empfohlen für:** 2./4. Semester

**Verantwortlich** Professur für Data Privacy and Security

**Dauer** 1 Semester

**Modulturnus** jedes Sommersemester

**Lehrformen**

- Vorlesung "Datenschutz von Anonymisierung bis Zweckbindung" (2 SWS) = 30 h Präsenzzeit und 60 h Selbststudium = 90 h
- Übung "Anwendung von Datenschutzverfahren" (1 SWS) = 15 h Präsenzzeit und 45 h Selbststudium = 60 h
- Seminar "Aktuelle Datenschutzverfahren aus der Forschung" (2 SWS) = 30 h Präsenzzeit und 120 h Selbststudium = 150 h

**Arbeitsaufwand** 10 LP = 300 Arbeitsstunden (Workload)

**Verwendbarkeit**

- M.Sc. Data Science: Bereich Skalierbares Datenmanagement
- M.Sc. Informatik: Vertiefungsmodul

**Ziele**

Nach der aktiven Teilnahme am Modul "Datenschutz von Anonymisierung bis Zweckbindung" verfügen die Studierenden über ein umfassendes Verständnis der gesellschaftlichen, technischen und wirtschaftlichen Ziele des Datenschutzes. Sie erwerben detaillierte theoretische und praktische Kenntnisse von grundlegenden Anonymisierungsverfahren und Anonymitätsmaßen. Dazu gehört auch die Abgrenzung der wesentlichen Fachbegriffe des Datenschutzes, insbesondere von denen aus der Datensicherheit. Die Studierenden können Datenschutzlösungen auf unterschiedlichen technischen und organisatorischen Ebenen anwenden und bewerten, sowie Systeme nach dem Standard-Datenschutzmodell absichern. Sie können technische Datenschutzlösungen im Internet zielführend einsetzen, z.B. Mix-Kaskaden, TOR, Freenet oder DoNotTrack.

Im Seminarteil erwerben die Studierenden die Fähigkeit, sich in komplexe neue Datenschutzmaße und -verfahren einzuarbeiten. Sie können Konzepte aus der Forschung kritisch analysieren und bewerten. Die Studierenden können diese Konzepte Fachfremden verständlich und überzeugend präsentieren und Alternativansätze gegenüberstellen.

**Inhalt**

Die Vorlesung führt in die gesellschaftlichen Grundlagen und Zielstellungen des Datenschutzes ein. Sie bietet einen breit angelegten Überblick über technische Lösungen des Datenschutzes mit Fokus auf Datenschutz im Internet. Zu diesen Lösungen zählen Anonymisierungsverfahren, Verfahren zum Verbergen der IP-Adresse oder der Urheberschaft von Informationen, sowie Ansätze zum Datenselbstschutz. In der Übung werden diese Konzepte an praktischen Beispielen vertieft.

Im Seminar werden die Studierenden vor die Aufgabe gestellt, für sie neue, komplexe Datenschutzprobleme und Datenschutzlösungen zu verstehen und zu bewerten, und diese so aufzubereiten und zu präsentieren, dass sie für Personen ohne Datenschutz-Fachwissen verständlich werden.

**Teilnahmevoraussetzungen** Teilnahme am Modul "Betriebs- und Kommunikationssysteme" (10-201-2004) oder gleichwertige Kenntnisse.  
Die Module 10-INF-DS105 und 10-INF-DS107 schließen sich gegenseitig aus.

**Literaturangabe** Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten** Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben.  
Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung:</b>	
Klausur 60 Min., mit Wichtung: 1	Vorlesung "Datenschutz von Anonymisierung bis Zweckbindung" (2SWS)
Referat 20 Min., mit Wichtung: 1	Übung "Anwendung von Datenschutzverfahren" (1SWS)
	Seminar "Aktuelle Datenschutzverfahren aus der Forschung" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-INF-DS202	Wahlpflicht

### Modultitel **Probabilistisches Maschinelles Lernen**

**Modultitel (englisch)** Probabilistic Machine Learning

**Empfohlen für:** 2. Semester

**Verantwortlich** Humboldt Professur für Künstliche Intelligenz

**Dauer** 1 Semester

**Modulturnus** jedes Sommersemester

**Lehrformen**

- Vorlesung "Probabilistic Machine Learning" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h
- Seminar "Probabilistic Machine Learning" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h

**Arbeitsaufwand** 5 LP = 150 Arbeitsstunden (Workload)

**Verwendbarkeit** • M.Sc. Data Science: Bereich Datenanalyse

**Ziele**

Nach der aktiven Teilnahme am sind die Studierenden in der Lage

- die Definitionen und Konzepte hinter Standardwerkzeugen für maschinelles Lernen zu verstehen
- Standardalgorithmen für maschinelles Lernen auf Daten anzuwenden
- die Programmiersprache Python und Jupyter zur Datenanalyse und zur Anpassung von Algorithmen für maschinelles Lernen zu verwenden
- die Berechnungen, Mathematik und Statistik, die grundlegenden Algorithmen für maschinelles Lernen zugrunde liegen, zu verstehen

After actively taking part in the course, the students should

- Understand the definitions and concepts behind standard machine learning tools
- Be able to apply standard machine learning algorithms to data
- Be able to use Python and Jupyter notebooks for data analysis and implementation of machine learning algorithms
- Understand the computations, mathematics, and statistics underlying basic machine learning algorithms

**Inhalt**

Regressionsmodelle

- Lineare Regression: Frequentistisch und Bayesianisch
- Nichtlineare Regression: Frequentistisch und Bayesianisch

Grafische Modelle

- Mixtures of Gaussians, Latent Dirichlet Allocation
- Markov chain Monte Carlo and Expectation Maximization

Tiefe Neuronale Netze

- Variations-Autoencoder und Dimensionsreduktion
- Stochastik Gradientenabstieg und Optimierung im großen Maßstab
- Automatisierte Differenzierung
- Architekturen neuronaler Netze

Das Modul wird in englischer Sprache gehalten. Lehr- und Prüfungssprache ist englisch.

Regression models:

- Linear regression -- Frequentist and Bayesian
- Nonlinear regression -- Frequentist and Bayesian

Graphical models

- Mixtures of Gaussians and Latent Dirichlet Allocation
- Markov chain Monte Carlo and Expectation Maximization

Deep Neural Networks

- Variational autoencoders and dimension reduction
- Stochastic gradient descent and large-scale optimization
- Automated differentiation
- Architectures of neural networks

**Teilnahmevoraussetzungen**

Kenntnisse in Linearer Algebra und Grundlegende Wahrscheinlichkeit

**Literaturangabe**

Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Klausur 60 Min., mit Wichtung: 1</b>	
<i>Prüfungsvorleistung: Projektarbeit (8 Wochen) mit schriftlicher Ausarbeitung</i>	
	Vorlesung "Probabilistic Machine Learning" (2SWS)
	Seminar "Probabilistic Machine Learning" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-INF-DS301	Wahlpflicht

### Modultitel **Aktuelle Trends in Data Science**

**Modultitel (englisch)** Current Trends in Data Science

**Empfohlen für:** 2. Semester

**Verantwortlich** Leitung des Instituts für Informatik

**Dauer** 1 Semester

**Modulturnus** jedes Sommersemester

**Lehrformen**

- Vorlesung "Aktuelle Trends in Data Science" (2 SWS) = 30 h Präsenzzeit und 70 h Selbststudium = 100 h
- Übung "Aktuelle Trends in Data Science" (1 SWS) = 15 h Präsenzzeit und 35 h Selbststudium = 50 h

**Arbeitsaufwand** 5 LP = 150 Arbeitsstunden (Workload)

**Verwendbarkeit** • M.Sc. Data Science

**Ziele** Nach der aktiven Teilnahme am Modul "Aktuelle Trends in Data Science" sind die Studierenden in der Lage, Begriffe und Verfahren eines aktuellen Themas im Bereich Data Science zu benennen und zu erklären.

Sie können ausgewählte Verfahren und Algorithmen analysieren, beurteilen und diese selbstständig auf Problemstellungen anwenden.

Die Studierenden sind imstande, verschiedene Ansätze in der Gruppe zu diskutieren.

**Inhalt** Wechselnde aktuelle Themen und Trends im Bereich Data Science, die sich besonderer Aktualität erfreuen. Die Veranstaltung kann auch von Gästen des Instituts für ein eigenes Lehrangebot genutzt werden, das nicht in die bestehenden Module passt. Der konkrete Inhalt wird im Vorlesungsverzeichnis bekannt gegeben.

**Teilnahmevoraussetzungen** keine

**Literaturangabe** Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten** Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

### Prüfungsleistungen und -vorleistungen

Modulprüfung: Mündliche Prüfung 30 Min., mit Wichtung: 1	
	Vorlesung "Aktuelle Trends in Data Science" (2SWS)
	Übung "Aktuelle Trends in Data Science" (1SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-201-2219S	Wahlpflicht

### Modultitel Grundlagen der Parallelverarbeitung (S)

**Modultitel (englisch)** Foundations of Parallel Processing (S)

**Empfohlen für:** 3. Semester

**Verantwortlich** Professur für Schwarmintelligenz und Komplexe Systeme

**Dauer** 1 Semester

**Modulturnus** jedes Wintersemester

**Lehrformen**

- Vorlesung "Grundlagen der Parallelverarbeitung" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h
- Seminar "Grundlagen der Parallelverarbeitung" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h

**Arbeitsaufwand** 5 LP = 150 Arbeitsstunden (Workload)

**Verwendbarkeit**

- M.Sc. Bioinformatik
- M.Sc. Data Science
- Lehramt Informatik

**Ziele** Nach der aktiven Teilnahme am Modul "Grundlagen der Parallelverarbeitung" sind die Studierenden in Lage:

- grundlegende Begriffe und Konzepte der Parallelverarbeitung zu formulieren und zu erklären,
- grundlegende parallele algorithmische Verfahren und Rechnermodelle (u.a. Sortieralgorithmen, Hardware- Addition) zu analysieren und zu vergleichen und
- für grundlegende algorithmische Probleme selbständig parallele Lösungsverfahren zu entwerfen.

**Inhalt** Es wird eine der folgenden Vorlesungen und das Seminar gewählt.

Vorlesung "Parallele Algorithmen": Grundlegende Konzepte und Bewertungskriterien für parallele Algorithmen, PRAM-Modell, Parallele Algorithmen für grundlegende Probleme wie Sortieren oder Mergen, Grundlagen von Hardware Algorithmen.

Vorlesung "Parallele Berechnungsmodelle": Grundlegender Aufbau von Parallelrechnern, Einführung in realistische Parallelerrechnermodelle, Varianten des BSP-Modells, Varianten des LogP-Modells, Auswirkungen der Modelle auf den Entwurf von Algorithmen, Algorithmische Lösung von Beispielproblemen.

Vorlesung "Rekonfigurierbare Rechensysteme": Einsatzbereiche rekonfigurierbarer Rechensysteme, Typen rekonfigurierbarer Rechensysteme, Aufbau von Field Programmable Gate Arrays (FPGAs), Theoretische Konzepte der Rekonfigurierbarkeit, Grundlegende Algorithmen zu dynamischer Rekonfiguration

Vorlesung "Entwurf und Implementierung paralleler Algorithmen": Parallele Plattformen, Entwurfsprinzipien, Analytische Modellierung, Parallele Programmierung für nachrichtengekoppelte und speichergekoppelte

Parallelrechner, Matrixmultiplikation, Sortieren, Graphenalgorithmen, Diskrete Optimierung, Dynamische Programmierung.

Seminar "Grundlagen der Parallelverarbeitung": Das Seminar behandelt ergänzende Arbeiten zu den Themen der Vorlesungen und befasst sich mit aktuellen wissenschaftlichen Abhandlungen aus der Parallelverarbeitung.

**Teilnahmevoraussetzungen**

Teilnahme am Modul "Algorithmen und Datenstrukturen 1" (10-201-2001-1) oder gleichwertige Kenntnisse.  
Die Module 10-201-2219S und -2219V schließen sich gegenseitig aus.

**Literaturangabe**

Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Mündliche Prüfung 20 Min., mit Wichtung: 1</b>	
	Vorlesung "Grundlagen der Parallelverarbeitung" (2SWS)
Referat (45 Min.) mit schriftlicher Ausarbeitung (2 Wochen), mit Wichtung: 1	Seminar "Grundlagen der Parallelverarbeitung" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-201-2219V	Wahlpflicht

### Modultitel Grundlagen der Parallelverarbeitung (V)

**Modultitel (englisch)** Foundations of Parallel Processing (V)

**Empfohlen für:** 3. Semester

**Verantwortlich** Professur für Schwarmintelligenz und Komplexe Systeme

**Dauer** 1 Semester

**Modulturnus** jedes Wintersemester

**Lehrformen**

- Vorlesung "Grundlagen der Parallelverarbeitung" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h
- Vorlesung mit seminaristischem Anteil "Grundlagen der Parallelverarbeitung 2" (2 SWS) = 30 h Präsenzzeit und 45 h Selbststudium = 75 h

**Arbeitsaufwand** 5 LP = 150 Arbeitsstunden (Workload)

**Verwendbarkeit**

- M.Sc. Bioinformatik
- M.Sc. Data Science
- Lehramt Informatik

**Ziele** Nach der aktiven Teilnahme am Modul „Grundlagen der Parallelverarbeitung“ sind die Studierenden in Lage:

- grundlegende Begriffe und Konzepte der Parallelverarbeitung zu formulieren und zu erklären,
- grundlegende parallele algorithmische Verfahren und Rechnermodelle (u.a. Sortieralgorithmen, Hardware- Addition) zu analysieren und zu vergleichen und
- für grundlegende algorithmische Probleme selbständig parallele Lösungsverfahren zu entwerfen.

**Inhalt** Studierende wählen die grundlegende Vorlesung "Parallele Algorithmen" und eine weiterführende Vorlesung.

Parallele Algorithmen: Grundlegende Konzepte und Bewertungskriterien für parallele Algorithmen, PRAM-Modell, Parallele Algorithmen für grundlegende Probleme wie Sortieren oder Mergen, Grundlagen von Hardware Algorithmen.

Weiterführende Vorlesungen

1. Parallele Berechnungsmodelle: Grundlegender Aufbau von Parallelrechnern, Einführung in realistische Parallelerrechnermodelle, Varianten des BSP-Modells, Varianten des LogP-Modells, Auswirkungen der Modelle auf den Entwurf von Algorithmen, Algorithmische Lösung von Beispielproblemen.

2. Rekonfigurierbare Rechensysteme: Einsatzbereiche rekonfigurierbarer Rechensysteme, Typen rekonfigurierbarer Rechensysteme, Aufbau von Field Programmable Gate Arrays (FPGAs), Theoretische Konzepte der Rekonfigurierbarkeit, Grundlegende Algorithmen zu dynamischer Rekonfiguration

In unregelmäßigen Abständen wird die grundlegende Vorlesung durch die folgende Vorlesung ersetzt:

Entwurf und Implementierung paralleler Algorithmen: Parallele Plattformen, Entwurfsprinzipien, Analytische Modellierung, Parallele Programmierung für nachrichtengekoppelte und speichergekoppelte Parallelrechner, Matrixmultiplikation, Sortieren, Graphenalgorithmen, Diskrete Optimierung, Dynamische Programmierung.

**Teilnahmevoraussetzungen**

Teilnahme am Modul "Algorithmen und Datenstrukturen 1" (10-201-2001-1) oder gleichwertige Kenntnisse.  
Die Module 10-201-2219S und -2219V schließen sich gegenseitig aus.

**Literaturangabe**

Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten**

Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Mündliche Prüfung 20 Min., mit Wichtung: 1</b>	
	Vorlesung "Grundlagen der Parallelverarbeitung" (2SWS)
	Vorlesung mit seminaristischem Anteil "Grundlagen der Parallelverarbeitung 2" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-INF-DS103	Wahlpflicht

### Modultitel **Praktikum für Data Warehousing und Data Mining**

**Modultitel (englisch)** Data Warehousing and Data Mining Project

**Empfohlen für:** 3. Semester

**Verantwortlich** Professur für Informatik (Datenbanken)

**Dauer** 1 Semester

**Modulturnus** jedes Wintersemester

**Lehrformen** • Praktikum "Praktikum für Data Warehousing und Data Mining" (2 SWS) = 30 h Präsenzzeit und 120 h Selbststudium = 150 h

**Arbeitsaufwand** 5 LP = 150 Arbeitsstunden (Workload)

**Verwendbarkeit** • M.Sc. Data Science

**Ziele** Nach der aktiven Teilnahme am Modul "Praktikum für Data Warehousing und Data Mining" sind die Studierenden in der Lage, aktuelle Datenbank-/Big-Data-Technologien anzuwenden.

Sie können selbständig Anfragen formulieren und ausführen.

Die Studierenden sind imstande, Integrationsverfahren sowie Techniken zur Anfragebearbeitung und -optimierung durchzuführen.

Sie sind in der Lage, Algorithmen zur effizienten Verarbeitung großer Datenmengen anzuwenden bzw. zu implementieren.

**Inhalt**

- Praktische Realisierung eines Data-Warehousing- und Data-Mining-Projektes
- Datenimport und -bereinigung (Objektkonsolidierung, Normalisierung, ...)
- Definition und Erstellung eines Data Cubes und OLAP-Analysen
- Anwendung und Implementierung verschiedener Data Mining Algorithmen unter Verwendung von existierenden Big Data Frameworks und Technologien

Je nach Interesse des Studierenden kann der Schwerpunkt des Praktikums entweder auf Data Warehousing oder auf Data Mining gelegt werden.

**Teilnahmevoraussetzungen** keine

**Literaturangabe** Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten** Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

**Prüfungsleistungen und -vorleistungen**

<b>Modulprüfung: Praktikumsleistung (3 Testate a 60 Min.), mit Wichtung: 1</b>	
	Praktikum "Praktikum für Data Warehousing und Data Mining" (2SWS)

## Master of Science Data Science

Akademischer Grad	Modulnummer	Modulform
Master of Science	10-INF-DS02	Pflicht

### Modultitel **Masterseminar Data Science**

**Modultitel (englisch)** Master's Seminar: Data Science

**Empfohlen für:** 4. Semester

**Verantwortlich** Studiengangsverantwortlicher Master of Science Data Science

**Dauer** 1 Semester

**Modulturnus** jedes Semester

**Lehrformen** • Seminar "Masterseminar Data Science" (2 SWS) = 30 h Präsenzzeit und 120 h Selbststudium = 150 h

**Arbeitsaufwand** 5 LP = 150 Arbeitsstunden (Workload)

**Verwendbarkeit** • M.Sc. Data Science

**Ziele** Nach der aktiven Teilnahme am Modul "Masterseminar Data Science" sind die Studierenden in der Lage sich selbstständig in aktuelle Forschungsthemen der Data Science einzuarbeiten.

Sie können selbstständig die zugehörige Literatur auswählen und analysieren, sowie das Thema angemessen in einer Gruppe präsentieren.

**Inhalt** In jedem Semester bieten mehrere Abteilungen Absolventenseminar an, das im Rahmen des Masterseminars belegt werden kann. Die Seminare behandeln aktuelle Forschungsthemen aus einem Forschungsgebiet der jeweiligen Abteilung.

**Teilnahmevoraussetzungen** keine

**Literaturangabe** Hinweise zu Literaturangaben erfolgen in den Lehrveranstaltungen.

**Vergabe von Leistungspunkten** Leistungspunkte werden mit erfolgreichem Abschluss des Moduls vergeben. Näheres regelt die Prüfungsordnung.

### Prüfungsleistungen und -vorleistungen

<b>Modulprüfung: Referat 60 Min., mit Wichtung: 1</b>	
	Seminar "Masterseminar Data Science" (2SWS)