



Modulbeschreibung

Advanced Information Management 3 (AIM-3) Scalable Data Science: Systems & Methods (SDSSM)

Modultitel: Advanced Information Management 3 (AIM-3) Scalable Data Science: Systems & Methods (SDSSM) Advanced Information Management 3 (AIM-3) Scalable Data Science: Systems & Methods (SDSSM)	Leistungspunkte: 6	Modulverantwortlicher: Markl, Volker
URL: http://www.dima.tu-berlin.de/	Sekretariat: EN 7	Ansprechpartner: Soto, Juan
	Modulsprache: Englisch	Kontakt: sekr@dima.cs.tu-berlin.de

Lernergebnisse

The last decade was marked by the digitalization of virtually all aspects of modern society. Today, businesses, government institutions, as well as science and engineering organizations, among others face an avalanche of digital data on a daily basis. In order to derive insight from all of this data, society needs individuals with a strong foundation in scalable data science. In this course students will learn about popular scalable data analysis systems and scalable data analytics methods and gain practical experience in conducting scalable data science.

no translation

Lehrinhalte

The module will focus on mainstream distributed processing platforms and paradigms and learn how to employ these to solve challenging big data problems using popular data mining methods. Students will learn how to implement and employ varying data mining algorithms, such as Naïve Bayes, K-Means Clustering, and PageRank on varying open-source systems (e.g., Apache Hadoop, Apache Flink).

no translation

Modulbestandteile

Lehrveranstaltungen	Art	Nummer	Turnus	SWS
Advanced Information Management 3 (AIM-3) - Scalable Data Science: Systems & Methods (SDSSM)	IV	0434 L 472	WS/SS	4

Arbeitsaufwand und Leistungspunkte

Advanced Information Management 3 (AIM-3) - Scalable Data Science: Systems & Methods (SDSSM) (Integrierte Veranstaltung)	Multiplikator:	Stunden:	Gesamt:
Exercises/Practice	15.0	4.0h	60.0h
Plenary sessions	15.0	4.0h	60.0h
Preparation & Consolidation (incl. literature studies)	15.0	4.0h	60.0h
			180.0h

Ein Leistungspunkt entspricht 30.0 Stunden (Es wird folgende Rundungsart verwendet: Aufrunden)

Beschreibung der Lehr- und Lernformen

This Integrated Course (Integrierte Veranstaltung, IV) consists of: (i) lectures on key concepts, (ii) practical theoretical & programming exercises, and (iii) student lead presentations (including literature search). Active participation and contributions to all parts of this course are essential.

Voraussetzungen für die Teilnahme / Prüfung

Wünschenswerte Voraussetzungen für die Teilnahme zu den Lehrveranstaltungen:

Computer science topics addressed in TU Berlin modules in the Bachelor's curriculum, particularly, the database course ("Information Systems and Data Analysis") or the equivalent, as well as good Java programming skills are required. Basic knowledge in linear algebra, numerical analysis, probability, and statistics are strongly recommended. Furthermore, it is preferable if students have already completed (or are currently enrolled in) a machine-learning course. Since the course will be offered in English, fluency in English is also required.

Verpflichtende Voraussetzungen für die Modulprüfungsanmeldung:

keine Angabe

Abschluss des Moduls

Prüfungsform:
Portfolioprüfung

Benotet:
benotet

The portfolio exam (worth 100 points) is comprised of three parts, namely: (i) written homework (30 points), (ii) in-class presentations (20 portfolio points), and (iii) a written exam (50 portfolio points).

The final grade according to § 47 (2) AllgStuPO will be calculated with the faculty grading table 2.
(Die Gesamtnote gemäß § 47 (2) AllgStuPO wird nach dem Notenschlüssel 2 der Fakultät IV ermittelt.)

Prüfungselement	Gewicht
(Deliverable assessment) Homework	30
(Deliverable assessment) in-class presentations	20
(Examination) Written test	50

Dauer des Moduls

Das Modul kann in 1 Semester(n) abgeschlossen werden.

Maximale teilnehmende Personen

Das Modul ist auf 30 Teilnehmer begrenzt.

Anmeldeformalitäten

Students are required to register via the DIMA course registration tool before the start of the first lecture (<http://www.dima.tu-berlin.de/>). Within the first six weeks after commencement of the lecture, students will have to register for the course at QISPOS (university examination protocol tool) and ISIS (course organization tool) in addition to the registration at the DIMA course registration tool.

Literaturhinweise, Skripte

Skript in Papierform:
nicht verfügbar

Elektronisches Skript:
Es wird ein elektronisches Skript angeboten

Hinweis zum elektronischen Skript:
<https://www.dima.tu-berlin.de/menue/teaching/>

Empfohlene Literatur:

Anand Rajaraman, Jeffrey David Ullman : Mining of Massive Datasets (Free Online: <http://infolab.stanford.edu/~ullman/mmds/book.pdf>)
Data Mining: Practical Machine Learning Tools and Techniques, Ian H. Witten and Eibe Frank, Morgan Kaufmann, 2011.
Hadoop: The Definitive Guide (4th Edition), Tom White, O'Reilly Media, 2015.
Supplementary reading material may be assigned to complement course lectures.

Zugeordnete Studiengänge

Die Modulversion wird auf keiner Modulliste verwendet.

This course targets Master's students focused on Database Systems and Information Management in Computer Science (Major: System Engineering), Computer Engineering (Major: Information Systems & Software Engineering), and Industrial Engineering. Compulsory Elective module for ERASMUS MUNDUS IT4BI, plus Compulsory for EIT-ICT Data Science (DS) and Compulsory Elective for EIT-ICT Cloud Computing and Services (CCS)
Subject to space availability, Master's students in other academic programs may also enroll and satisfy elective module requirements.

Wahlpflichtmodul im Masterstudiengang Informatik/Studienschwerpunkt System Engineering, Tech-nische Informatik/Studienschwerpunkte Informationssysteme & Software Engineering und im Master-studiengang Wirtschaftsingenieurwesen (Studiengang IuK). Wahlpflichtmodul im ERASMUS MUNDUS IT4BI, sowie für EIT-ICT Cloud Computing and Services (CCS), Pflicht für EIT-ICT Data Science (DS).
Je nach Verfügbarkeit der Plätze können auch Studierende anderer Fachrichtungen als Wahlpflicht das Modul belegen.

Sonstiges

Since 2014, this module is offered each summer and winter term.
For each topic during this course additional research papers and reports will be used.