



Cursusomvang (nominale waarden; effectieve waarden kunnen verschillen per opleiding)

Studiepunten 6.0 Studietijd 165 u Contacturen 62.5 u

Aanbodsessies en werkvormen in academiejaar 2018-2019

A (semester 2)	Engels	begeleide zelfstudie	15.0 u
		werkcollege: PC- klasoefeningen	15.0 u
		zelfstandig werk	10.0 u
		hoorcollege	22.5 u

Lesgevers in academiejaar 2018-2019

Saeyns, Yvan	WE02	Verantwoordelijk lesgever
Goetghebeur, Els	WE02	Medelesgever
Ley, Christophe	WE02	Medelesgever

Aangeboden in onderstaande opleidingen in 2018-2019

	stptn	aanbodsessie
Master of Science in de informatica	6	A

Onderwijstalen

Engels

Trefwoorden

Big data, Map-Reduce, large-scale machine learning, graph mining

Situering

Familiarize the students with the most important methods to extract information from large datasets using modern data mining techniques. The students are expected to learn how to use these techniques correctly in applications and they acquire the skills to interpret obtained results in a statistically correct manner.

The students will also be introduced to big data and the problems this might impose.

Inhoud

Distributed File Systems

- the Google File System
- HDFS

The Map-Reduce Framework and its extensions

- Map-Reduce basics
- Large-scale matrix computations
- Large-scale database operations
- SPARK

Data structures for big data settings

- Mining large graphs
- Locality sensitive hashing (LSH)

Data Visualization

- Efficient data visualization
- Dealing with large and high-dimensional data

No-SQL databases

Itemset mining

Mining streaming data

Case studies in big data

Begincompetenties

This course requires the following background: good knowledge of data structures, programming, basics of machine learning and parallel computing.

Eindcompetenties

- 1 Explain the techniques used for data fragmentation, replication, and allocation during the distributed database design process. [Familiarity]
- 2 Evaluate simple strategies for executing a distributed query to select the strategy that minimizes the amount of data transfer. [Assessment]
- 3 Compare and contrast different uses of data mining as evidenced in both research and application. [Assessment]
- 4 Explain the value of finding associations in market basket data. [Familiarity]
- 5 Characterize the kinds of patterns that can be discovered by association rule mining. [Assessment]
- 6 Describe how to extend a relational system to find patterns using association rules. [Familiarity]
- 7 Evaluate different methodologies for effective application of data mining. [Assessment]
- 8 Identify and characterize sources of noise, redundancy, and outliers in presented data. [Assessment]
- 9 Identify mechanisms (on-line aggregation, anytime behavior, interactive visualization) to close the loop in the data mining process. [Familiarity]

Creditcontractvoorwaarde

Toelating tot dit opleidingsonderdeel via creditcontract is mogelijk mits gunstige beoordeling van de competenties

Examencontractvoorwaarde

Dit opleidingsonderdeel kan niet via examencontract gevolgd worden

Didactische werkvormen

Begeleide zelfstudie, hoorcollege, zelfstandig werk, werkcollege: PC-klasoefeningen

Toelichtingen bij de didactische werkvormen

Minerva will be used to ensure a smooth organisation and follow-up of the practical assignments.

Leermateriaal

This course uses the following handbook:

"Mining of massive datasets" Jure Leskovec, Anand Rajaraman and Jeffrey D. Ullman
The pdf of this book is available free of charge. Additional slides are available on Minerva.

Referenties

Mining of massive datasets (Leskovec, Rajaraman and Ullman)

Vakinhoudelijke studiebegeleiding

The exercises and practical assignments are supervised by the lecturer.

Evaluatiemomenten

periodegebonden evaluatie

Evaluatievormen bij periodegebonden evaluatie in de eerste examenperiode

Mondeling examen, werkstuk

Evaluatievormen bij periodegebonden evaluatie in de tweede examenperiode

Mondeling examen, werkstuk

Evaluatievormen bij niet-periodegebonden evaluatie

Tweede examenkans in geval van niet-periodegebonden evaluatie

Niet van toepassing

Toelichtingen bij de evaluatievormen

Project work shows the student's capability to analyse real data problems by using the methods taught in this course and interpret correctly the obtained results.

Oral examination based on a written report about the results of the student's project work.

Eindscoreberekening

100% examination