Analysis of information systems for CSE

Analysis of information systems for CSE	
Offered by	Department of Mathematics and Computer Science
Language	English
Primarily interesting for	BSc in Computer Science and Engineering (CSE)
Prerequisites	Students are assumed to have basic skills in logic, set theory, calculus, discrete mathematics, databases, algorithms and programming
Contact person	dr. O. Papapetrou, o.papapetrou@tue.nl

Content and composition

Provide a description of the content and composition of the elective package.

Course code	Course name	Level classification
2ID70	Data-intensive systems and applications	3
JBI100	Visualization	2
2IX30	Responsible data science	3

Course description

2ID70, Data-intensive systems and applications
This course prepares students to meet the new challenges of contemporary data engineering in which traditional assumptions break, where new data models, query languages and programming interfaces are required. In this course, we study how traditional relational database techniques such as indexing, query planning and optimization, transaction management and self-tuning can be made to work on a massive scale of thousands of machines and petabytes of data. We study models of contemporary data-intensive systems, their efficient engineering, and their practical use. These models include scalable data processing platforms (e.g., MapReduce, Spark) and stream processing engines. We discuss why these models were introduced, their relative advantages and disadvantages, how they are engineered, and how to effectively use them in practice.

JBI100, Visualization
In the visualization course you will learn about the challenges of visually representing data that comes in a variety of forms. Starting from simple primitive data types like categorical, ordinal, or quantitative data, we will have a look into more complex dataset scenarios including relational data like graphs/networks or hierarchies, multivariate data, text data, or trajectory data that contains an inherent spatio-temporal aspect. In this course you will learn about the data processing, data transformation, data visualization, and finally, the interaction with the visual output. To make a visualization interpretable, readable, and intuitive, we will also have a look at perceptual issues like pre-attentive processing, the visual memory, or Gestalt principles. Moreover, a number of laws or no-goes will be discussed to make the diagrams or visualization techniques more perceptually effective.

Analysis of information systems for CSE

2IX30, Responsible data science

The course is focused on studying the problems of fairness, accountability, confidentiality, and transparency (FACT) in data science, and data mining and machine learning in particular. One important challenge to face is that machine learnt models typically are not 100% accurate, i.e. in some ways these models are wrong. Thus, it is important to study how we can make a good use of models that are not perfect, how we can understand the strengths and weaknesses of these models, how we can help a decision maker to trust (or not trust) the model or its particular prediction, and how we can get insights into impact of input features and some inner logic of a predictive model. We need techniques not just to explain the decision of a model, but also to uncover and characterize undesired or even unlawful biases in its performance. Hence, the other important challenge to study is how to formally define such biases, how to uncover and quantify them and how to design machine learning solutions that would enable the so-called fair algorithmic decision making by design. On the other side of the spectrum, there are challenges of privacy and confidentiality. We will study the main principles and techniques that have been researched and employed in data mining for privacy preserving and secure computation to induce models from data and to apply them in real-life scenarios.