Analytics of information systems
BSc Elective Package

Offered by: Department of Mathematics and Computer Science
Language: English
Primarily interesting for: BSc in: Computer Science and Engineering (CSE), Industrial Engineering (IE), Applied Mathematics (AM), Data Science (DS)
Prerequisites: Students are assumed to have basic skills in logic, set theory, discrete mathematics, and programming.

Content and composition

Analysis of information, data, and knowledge is increasingly important, with broad application across science, engineering, society, and industry. To tackle these challenges, knowledge and skills in the management, mining, and analysis of (big) data collections is necessary. This elective package provides deeper study of the foundations and applications of analysis of data and information systems. Students should select a coherent package of courses from the following list.

<table>
<thead>
<tr>
<th>Course code</th>
<th>Course name</th>
<th>Scheduled (Quarter/Slot)</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>2ID70</td>
<td>Data-intensive systems and applications</td>
<td>Q3 / B</td>
<td>2</td>
</tr>
<tr>
<td>JBI100</td>
<td>Visualization</td>
<td>Q4 / A</td>
<td>2</td>
</tr>
<tr>
<td>2IID0</td>
<td>Web analytics</td>
<td>2018/19: Q2 / B</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2019/20: Q3 / C</td>
<td></td>
</tr>
<tr>
<td>2IOI0</td>
<td>DBL Process mining</td>
<td>Q3 / C</td>
<td>2</td>
</tr>
</tbody>
</table>

Based on your program of study, a coherent package is defined as follows:
- for CSE: If 2IOE0 is chosen as DBL in major, then package needs to contain 2ID70, JBI100 and 2IID0; else 2ID70, JBI100, and 2IOE0 or 2IID0
- for IE: 2ID70, JBI100, and 2IOE0 or 2IID0
- for AM: 2ID70, JBI100, and 2IOE0 or 2IID0
- for DS: 2ID70, 2IID0 and 2IOE0

Course descriptions
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 are among: Graph databases, document databases (NoSQL, JSON stores, etc.), and scalable data processing platforms (MapReduce, Spark, Flink, etc.). We discuss why these models were introduced, their relative advantages and disadvantages, how they are engineered, and how to effectively use them in practice.

March 2017
Analysis of Information Systems, BSc elective package
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.

Web analytics
During this course the students master the basics of intelligent data analysis, and obtain practical skills of being a data analyst, data miner, business analyst and experimenter by participating in practical exercises. In the course we focus on the utility of different technologies for understanding the Web as a whole, for modeling the behavior of the individual Web users and user groups, for using the Web as an experimentation platform. We will cover the topics of OLAP style data exploration, data mining, and social network analysis. We will give a special attention to understanding search, recommendations and advertisements on the Web.

DBL Process mining
In addition to data storage and data analysis, analysts need to relate data to process analysis. Process mining bridges the gap between traditional model-based process analysis (e.g., simulation and other business process management techniques) and data-centric analysis techniques such as machine learning and data mining. Process mining seeks the confrontation between event data (i.e., observed behavior) and process models. The goal of this course is to practice solving process mining problems, with a focus on hands-on skills in implementing and experimentally investigating their solutions. A second goal is to practice various professional skills, such as working effectively in a team, planning, technical writing and presenting.

Prior knowledge
The courses can be followed in any order, but need to take the specific prerequisites for each course into account when scheduling this package.