

TIETA17 Introduction to Big Data Processing Autumn 2016

- Instructor: university lecturer Heikki Hyyrö (heikki.hyyro@uta.fi).
- The 5 ECTS course "Introduction to Big Data Processing":
 - ♦ provides a practical introduction to the topic, with emphasis to the computing aspect.
 - ♦ Fairly light on theory.
 - Does not go into algorithmic details (e.g. of machine learning).
 - Deep knowledge of mathematics or statistics is not required.
 - ♦ However: the course may be challenging without satisfactory prior programming skills (and readiness to put effort into programming).
 - Exercises involve (mostly) Python and R programming languages and the Apache Hadoop and Spark distributed computing frameworks.

Lectures

- Lectures:
 - ♦ Wednesdays 31.8 – 19.10 at 14 – 16. (once per week, 8 weeks)
 - ♦ In the lecture room **LS A3**.
- Two lectures will have guest lecturers:
 - ♦ Representatives of the local IT sector / industry:
 - 28.9: Tapio Rautonen from Gofore (www.gofore.com).
 - 5.10: Mika Karaila from Valmet (www.valmet.com).
- Lecture material (slides etc.) and course exercise assignments will be published in the WETO-system.
 - ♦ <https://wetodev.sis.uta.fi/> (login with your university account)
- WETO will also serve as an electronic submission system for returning your answers to the exercises. The exercises will also be peer-reviewed: students give feedback to other students (also this is done in WETO).

Exercises

- The course involves 4 batches of mandatory exercise questions/tasks.
 - ♦ Questions are published in WETO.
 - ♦ Students submit their answers to WETO.
 - ♦ Student also peer-review some other student's answer(s) via WETO.
- The exercise schedule (batch publishing date → deadline for answers):
 - ♦ Batch 1: Friday 2.9 → Monday 12.9.
 - Basic data operations with Python.
 - ♦ Batch 2: Friday 16.9 → Monday 26.9.
 - Data gathering, transforming and/or analysis with R and Python.
 - ♦ Batch 3: Friday 30.9 → Tuesday 11.10.
 - Processing (big) data with Hadoop MapReduce.
 - ♦ Batch 4: Friday 14.10 → Monday 24.10.
 - Processing (big) data with Hadoop Spark.

Exercises...

- There are two exercise sessions per each exercise batch:
(1) a tutoring session and (2) a review session.
- The tutoring sessions: students may seek advice regarding the exercises.
 - ♦ Participation is voluntary (e.g. does not affect course grading).
- The review sessions: students give feedback about other students' answers and receive feedback from other students and/or the instructor.
 - ♦ Participation is **mandatory**: you receive points for the exercise batch questions only if you attend the corresponding review session.
 - Exceptions are allowed only for unavoidable reasons (work, illness etc.): notify the instructor beforehand if you cannot attend!
- All (but one exception) exercise session meetings are in Pinni B0016.
 - ♦ One-hour tutoring sessions meet on Wednesdays at 16-18.
 - ♦ Review sessions meet on Wednesdays at 16-18 and on Fridays at 14-16.
 - Attend a review session either on Wednesday or Friday, not both!

Exercise session meeting schedule...

- The schedule for exercise tutoring and review session meetings:
 - ♦ Exercise batch 1:
 - Tutoring session: Wednesday 7.9 at 16-18.
 - Review session: Wednesday 14.9 at 16-18 / Friday 16.9 at 14-16.
 - ♦ Exercise batch 2:
 - Tutoring session: Wednesday 21.9 at 16-18.
 - Review session: Wednesday 28.9 at 16-18 / Friday 30.9 at 14-16.
 - ♦ Exercise batch 3:
 - **Tutoring session: Monday 10.10 at 16-18 Pinni B0020. ([exception](#))**
 - Review session: Wednesday 12.10 at 16-18 / Friday 14.10 at 14-16.
 - ♦ Exercise batch 4:
 - Tutoring session: Wednesday 19.10 at 16-18.
 - Review session: Wednesday 26.10 at 16-18 / Friday 28.10 at 14-16.

Withdrawing (canceling course registration)

- According to the university policy, all registered students will receive a grade regardless of whether or not they pass the course.
 - ♦ The grade "**failed**" will be recorded to the student transcript of all registered students who (for any reason) fail to pass the course.
- If a student wishes to withdraw from this course, it must be done by Sunday 11.9 23:59 by withdrawing from the WETO course.
 - ♦ Further instructions are provided on the course WETO homepage.
- Those who remain as members of the WETO course after 11.9 can no longer cancel their registration; e.g. dropping out will result in "**failed**".

Course grading

- The course grade (1-5, or failed) will be primarily based on:
 - ♦ Exercises ($\approx 40\%$ of the grade).
 - Based on both the number (and quality) of answered exercise questions and the quality of feedback given to other students.
 - ♦ The final exam ($\approx 60\%$ of the grade).