FUNDAMENTALS OF PROGRAMMING

Week 10 – Data analytics, Dictionary, CSV, JSON and own modules

Janne Parkkila, Dr.Sc. (tech)

DATA ANALYTICS

Analysis paralysis?

WHAT IS DATA ANALYTICS?

▶▶ Data analytics means analyzing data

▶▶ Some consider it to be part of Artificial Intelligence

  ▶▶ (it is a requirement for AI, not really AI in itself)

▶▶ Where it fits?

  ▶▶ Data layer: Databases, APIs, .csv-files

  ▶▶ Analytics layer: Python, R, Matlab, Excel

  ▶▶ Visualization layer: Excel, matplotlib, Websites

  ▶▶ AI layer: Neural Networks, Machine Learning, (these are just algorithms)

DATA ANALYTICS

▸▸ Data analytics contains data evaluation, data cleaning, data transformation and modelling

▸▸ The goal is to find useful information for decisionmaking, either to support or automate

▸▸ There are many concepts, e.g. Data Science, Business Analytics

▸▸ We look at the programming part – how to process raw data and use it somehow

PROCESS PHASES

▶▶ 1. Data collection

   ▶▶ We need to gather our data somehow – Public open data, own diary, bank statements, bitcoin prices, tweets of Elon Musk

▶▶ 2. Data processing

   ▶▶ Process the data into usable format - tables, lists, objects => anything that is easy to process programmatically

▶▶ 3. Data cleanup

   ▶▶ Remove faulty data, fill in or filter out missing data, check data sensibility

▶▶ 4. Explorative data analysis

   ▶▶ Calculate statistical values, try to find interesting trends, understanding data

HANS ROSLING & DATA

▸▸ https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen?subtitle=en