Visualization II Seminar

Bára Kozlíková, Honza Byška, Vít Rusňák
& others

Spring 2020
Seminar 1

MUNI

VISITlab

Who we are?

- ## Bára Kozlíková

  - Head of visitlab, research in visualization, namely in biochemistry, criminology, geology, …

- ## Honza Byška

  - Visualization research, 4 years of PostDoc at University of Bergen

- ## Vít Rusňák

  - Researcher in the Proactive Security Group – Computer Security Incident Response Team at the Institute of Computer Science

- ## Tomáš Pšorn

  - PhD student in visitlab, employed at the Institute of Scientific Instruments of the Czech Academy of Sciences

VISITlab

What we will do here?

- Working on a visualization project …

- This project will be the only requirement for the credits and exam → select wisely ☺

VISITlab

How this will look like?

- We will give you a list of topics to choose from

- You can work on them individually or in groups

  - The number of people in the group will influence the required complexity of the final solution

- You will follow the workflow used in the visualization research

- The main goal is to **understand what it means to do a research in visualizations**

VISITlab

How to get the points?

- 5 tasks, for each you can get maximally 10 points

- Minimum for each task is 3 points to pass

- Minimum for passing the whole course is 26 points in total

50-46 points ………… A
45-41 points ………… B
40-36 points ………… C
35-31 points ………… D
30-26 points ………… E
25-0 points ………….. F

Tasks

- Tasks 1: Related work research, analysis of task requirements, parsing the data, preparing the sketch of the initial design
    10 points for initial sketches and parser

- Task 2: Finalizing the design, starting with implementation
    10 points for final design

- Task 3: Finalizing the implementation
    10 points for implementation

- Task 4: Presenting the result
    10 points for presentation

- Task 5: Writing the project report
    10 points for report

Before showing you the topics …

- Examples of the research procedure

VISITlab

Topic 1

Cellular Signal Coverage

The Czech Open Data portal contains various public datasets. Some of them describe cellular signal coverage on highways, railways, or via stationary measurements. We need to:

- show coverage by operators, or by technology (3G, 4G) on highways, motorways and from stationary measurements
- provide signal coverage analysis in the selected region/district
- identify blind spots or places with poor signal
- provide summary information for defined regions

https://data.gov.cz/datové-

sady?kl%C3%ADčová%20slova=mobiln%C3%AD%20s%C3%ADtě&kl%C3%ADčová%20slova=LTE&poskytovatel

=Český%20telekomunikačn%C3%AD%20úřad&velikost%20stránky=80

Topic 2

Marvel Studio open data

Marvel Studio provides API to dataset containing information about all fictional characters and comics of Marvel Universe they ever produced as well as their authors. This task focuses on graph and timeline visualizations showing relations among entities. Using proper visualization techniques, we can show:

- Relations among characters
- Clusters of characters (heroes vs. villains, Avengers vs. X-men)
- Timeline and important events of characters (related to comics)
- And many more...

https://developer.marvel.com

VISITlab

Topic 3

Junák – Czech scouting data

The dataset contains registration data of Junák - Czech scouting, for the last ten years. It includes not only the total number of registered persons but also their age category. The data visualization can give us:

- an insight into the long-term view and trends
- comparison of a single unit to the global statistics (i.e., growing slower/faster than the whole organization)
- comparison of multiple units of the same type
- the unit membership structure
- visualize the organization hierarchy
- overall statistics per year or multiple years
- approximate localization of units in the Czech Republic

https://is.skaut.cz/opendata/default.aspx

Topic 4

Antarctica measurements

This dataset contains the measurements of several time-dependent phenomena, captured at the Czech polar station in Antarctica. It contains the information about the evolution of snow level (captured by camera traps, monitoring the bamboo sticks), temperature of soil (measured by sensors in different depths), and wind speed and direction. When exploring the dataset, we need to:

- Understand the correlation between the measured phenomena
- Design an appropriate technique for the exploration the time-dependent data on different scales (months, days, hours, …) where simple aggregation over time is not possible

https://www.dropbox.com/sh/iaxhpqlv6vj3w1h/AADvVxGROCPQGsURPcrHDPHba?dl=0

Topic 5

Analysis of molecular dynamics

In this dataset, we are exploring properties of very large (50,000 frames) time series of molecular dynamics simulation. The data contains information about a protein, ligands, protein tunnel and water molecules. Similarly to Topic 4, when exploring this dataset, we need to:

- understand the correlation between the measured phenomena
- design an appropriate technique for aggregation large amount of data

https://www.dropbox.com/sh/4fog6mff2l7ca5b/AAA50IBzkNChtQLgChWsO5sQa?dl=0

# Topic 6

Technical inspection stations

Cesky Rozhlas created a dataset containing data from technical inspection stations (STK). The dataset has over 1 GB and includes reports from inspections realized in 2018. For each inspection, there is a vehicle identifier, vehicle vendor and type, mileage, fault categories, or inspection date.

We need to interpret stored data in multiple ways:

- Statistical information for each station (number of checked vehicles, reported faults, category of vehicles, passed inspection ratios, etc.) – per year, per month, per week
- Overall statistics – per year, per month, per week
- Vehicle statistics (average mileage, faults per vendor, per model, relations between age, mileage and number of faults, etc.)
- The interface should also enable basic search, filtering and sorting either by keyword search or by "lasso" selection in map
- As an extension, the information about the rating given by customers (e.g., from Google) can enhance the data.

https://data.gov.cz/datová-sada?iri=https%3A%2F%2Fdata.gov.cz%2Fzdroj%2Fdatové-sady%2Fhttps---un7pp4qfr5.execute-api.eu-west-1.amazonaws.com-prod-package_show-id-stk_md_2018

https://www.mdcr.cz/Dokumenty/Silnicni-doprava/STK/STK-Seznam-STK-dle-kraju

Topic 7
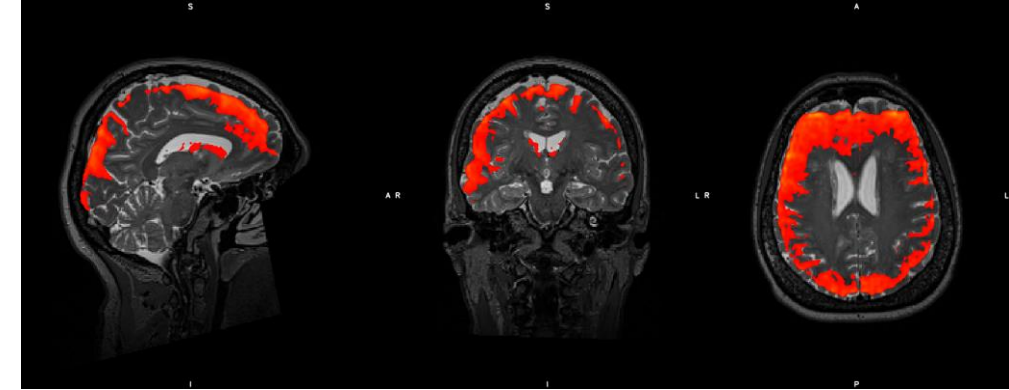
MR data visualization project

Medical image enhancement (changing brightness, contrast, mapping function) is a non-trivial task already in cases when one image is being enhanced. What about when you have two images and you want to enhance one of them? What if we want to alter opacity? Can we link enhancement and enhance two images simultaneously?

The task of this project is to create a web visualization engine for visualization of multidimensional MR data preferably using Django and VisPy package (you can suggest other frameworks). Furthermore, the aim is to explore the possibilities of image enhancement and answer some of the research questions.

So what now?

- Think about the project topics, have questions

- Discuss about that with your colleagues, for groups (if you want to)

- Start to think about technologies, check the data

- **Until next time…**

    - Final selection of the topic

    - Groups formed

VISITlab

Next seminar …

- Analysis of task requirements

- Choosing the technology

- Writing parser for data

- Starting to discuss the design