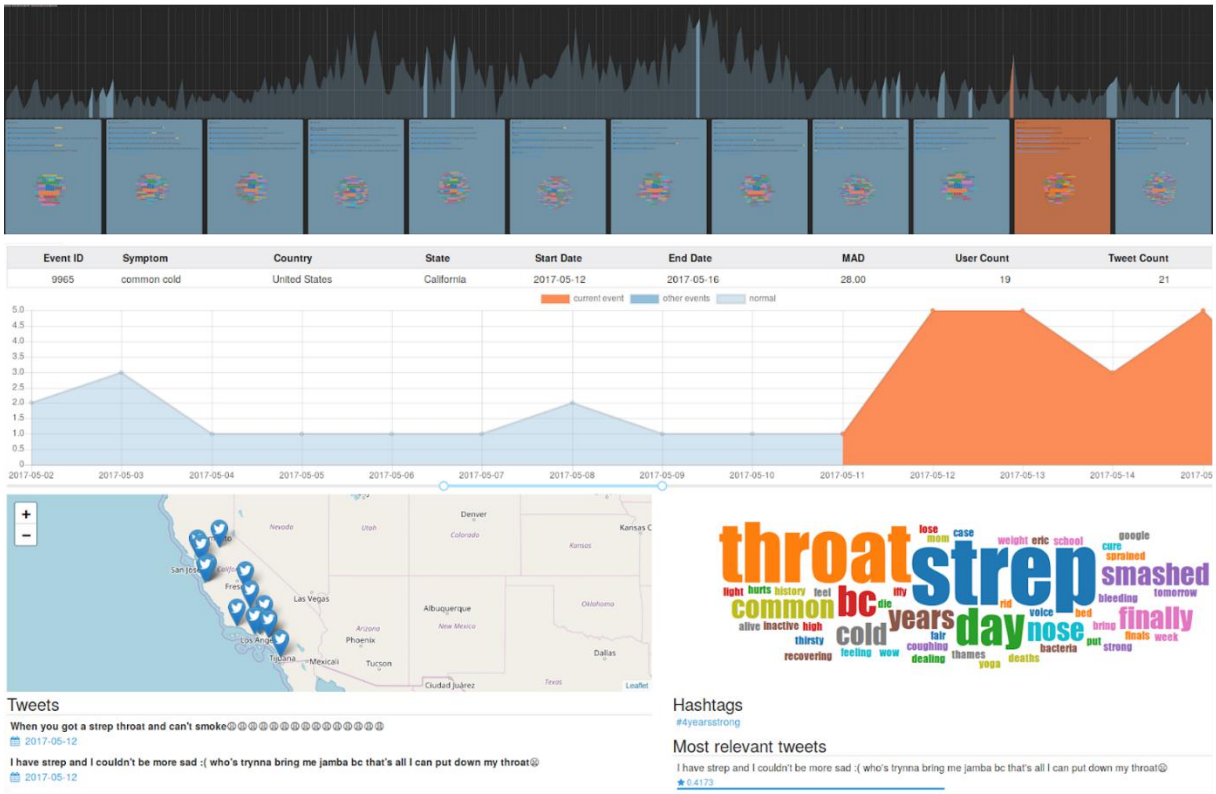
Imperial College London

University of the Year 2022 by The Times and Sunday Times' Good University Guide presents

Summer Online Programme

Data Curation for Interactive Visualisation

Learn from Imperial's Data Science Institute expert live online!



PROGRAMME OVERVIEW

Everyone starts to be aware that data is important, but society is still lacking skills and tools to understand large datasets. Humans and AI applications are producing more data than ever, so it becomes more important to assess the data quality and build good storytelling scenarios by using data visualisation techniques.

This programme will provide participants with an understanding of these technologies, apply the knowledge and learning experience to design, develop data curation and visualisation techniques specific to real-world datasets.

Topics covered include:

- **Data acquisition and quality**

This will be a hands-on guide on how to deal with data in the beginning of the processing pipeline and understand how the data quality has an impact on the overall results.

- **Data Processing and Curation**
  To introduce the idea of data curation and how is it different from a data processing pipeline.
- **Interactive Data Visualisation**
  To teach the basics in creating pragmatic interactive data visualisation by using modern web technologies.
- **Large scale data visualisation**
  To introduce the concepts of processing datasets that are large enough to not fit onto a single machine and using the Open Visualisation Environment to create a virtually infinite canvas to render your graphics.

Team based learning via group project:

As part of this programme, students will have the opportunity to work in small teams on a group project. Students will be asked first to select a novel dataset that has not been cleaned up. A list of sample datasets will be provided on the first week lecture, but students are also encouraged to bring their own data to work on. The group should be focusing on developing a project that uses some of the data curation, processing and visualisation techniques presented in the course work.

On completion of this programme, students will:

- Describe the latest development of data curation and processing methods
- Understand the basic knowledge about various types of data, storage, encoding and decoding techniques
- Apply the knowledge and experience gained to develop good data visualisation stories
- Be able to "debug" data quality issues by using various data curation and visualisation technique

PROGRAMME STRUCTURE AND TEACHING METHODS

- 5 live lectures delivering via Microsoft Teams with group discussions and questions.
- Web based exercises and quizzes will be provided for formative feedback.
- Group projects for assessing the learning outcomes, supported by tutorials.
- 1 session on final day for project presentation.

Project work will be done through team-based learning and tutorials. Final projects will be presented in groups on the last day of the programme. A prize will be awarded to the team with the best project.

The programme will be delivered over Microsoft Teams. Online project channels will be allocated to each team for project work. Students will be able to use the channel at any time to work on their project.

The entire programme will be taught in English.

CERTIFICATION

Students will receive a verified Imperial College London digital certificate on successful completion of this programme and a prize will be awarded to the best project team. Each student will also receive a transcript for their project marks.

ENTRY REQUIREMENTS

All students are expected to be studying an undergraduate or a postgraduate degree in a **technical subject**, i.e**.** Engineering, Computing, Software Engineering, Math, Physics or related disciplines, at a well-recognised university in China.

English requirements:

All students are required to have a good command of English, and if it is not their first language, they will need to satisfy the College requirement as follows:

- a minimum score of IELTS (Academic Test) 6.5 overall (with no less than 6.0 in any element) or equivalent.
- TOEFL (iBT) 92 overall (minimum 20 in all elements)
- CET- 4 (China) minimum score of 550
- CET- 6 (China) minimum score of 520

Technical requirements:

All students are expected to have a good level of programming skills, including basic principles of Python, JavaScript and databases (i.e. MongoDB, PostgreSQL)

Students will need to have access to a computer with a webcam, microphone and good internet connection to attend the live classes.

PROVISIONAL SCHEDULE

Session 1: The relation between data acquisition, curation, processing, and visualisation
Led by: Dr Ovidiu Șerban

Content

- Introduction to data types
- The relationship between different types of data, and curation techniques
- Applications of Data Visualisation
- Data Visualisation for story-telling skills
- Sample datasets to be used for the used for the group project
- Q&A

Session 2: Data quality and curation
Led by: Dr Ovidiu Șerban

Content

- What is data quality and how is it linked to data acquisition?
- Data Annotation and the relationship with data quality
- Data Curation
- Quick introduction into dataset storage and encoding
- Data Visualisation techniques applied to data curation and quality
- Q&A

Tutorial 1

Session 3: Working with numerical data
Led by: Dr Ovidiu Șerban

Content

- Storing and encoding numerical data
- Data curation for numerical data
- Data errors and imputation techniques
- Visualisation techniques and practices specific to numerical data
- Q&A

Session 4: Working with non-numerical data
Led by: Dr Ovidiu Șerban

Content

- Text data:
    - storing and encoding
    - curation and processing
- Other data types
    - Mapping data
    - Networks and graphs
- Visualisation techniques and practices specific to non-numerical data
- Q&A

Tutorial 2

Session 5: Interactive Large Scale Data Visualisation

Led by: Dr Ovidiu Șerban

Content

- Classic data visualisation techniques
- Interactive data visualisation
- Large-scale data visualisation with the Open Visualisation Environment (OVE)
- Q&A

Tutorial 3

Session 6: Project Presentations

Led by: Dr Ovidiu Șerban

Content

- Group project presentations
- Q&A and feedback
- Announcement of winning project group.

Programme ends.

THE PRESENTER

Dr Ovidiu Serban
Research Fellow in Intelligent Data Processing and Curation, Data Science Institute,
Imperial College London.
https://www.imperial.ac.uk/people/o.serban

Ovidiu Șerban is a Research Fellow at the Data Science Institute (DSI), Imperial College London. His current work includes real-time Natural Language Processing, Data Curation and Large Scale Visualisation Systems.

Ovidiu's research topics are Natural Language Processing, Machine Learning, Affective Computing and Interactive System Design. He holds a joint PhD from INSA de Rouen Normandy (France) and "Babeș-Bolyai" University (Romania), while working at LITIS Laboratory in France.

In his youth, Ovidiu worked at the Institute for Security Science and Technology (ISST), Imperial College London; Computer Lab, University of Cambridge, UK and ISR Laboratory, University of Reading, UK.

IMPERIAL COLLEGE LONDON AND THE DATA SCIENCE INSTITUTE

Consistently rated amongst the world's best universities (3rd in Europe and 7th in the World, QS World University Rankings 2022), Imperial College London is a science-based institution with an international reputation for excellence in teaching and research. Imperial attracts over 19,000 students and 8,000 staff of the highest international quality from over 136 different countries. Imperial has recently been named University of the Year 2022 by The Times and Sunday Times' Good University Guide.

Since its foundation in 1907, Imperial's contributions to society have included the discovery of penicillin, the development of holography and the foundations of fibre optics. This commitment to the application of research for the benefit of all continues today, with current areas of focus including interdisciplinary collaborations to improve global health, tackle climate change, develop sustainable sources of energy, address security challenges, develop data management and analysis technologies for supporting data driven research, and tackling problems at molecular scale.

Imperial's Centre for Continuing Professional Development had extensive experience in developing and running a range of online summer schools for undergraduate students. We draw on Imperial's education pedagogy in online learning to design and deliver summer schools that provide an engaging learning experience for students. Various interactive applications are used to support live teaching, online group projects are designed to assess students' learning outcomes and virtual social platform created in Flipgrip will provide students with a networking environment.

The Data Science Institute (DSI) is a major Imperial College London initiative that brings together Imperial's existing data science activities and expertise, and provides a focus and a catalyst for new partnerships.

The DSI supports multidisciplinary collaborations between the College's academic experts in many disciplines such as healthcare, financial services, climate science, and city infrastructure to create solutions to complex problems. Alongside research, the Institute fosters the next generation of data scientists and engineers by developing a range of postgraduate and executive courses.

The DSI includes 7 Academic Labs, has attracted over £50m in funding for data science research, technology and infrastructure and has published over 300 papers.

The Institute's Data Observatory (DO) was one of the first and largest visualisation suites in Europe. It provides a multi-dimensional and immersive environment to analyse large and complex data sets and to work collaboratively.

Thanks to its many research collaborations both across College and with a variety of external academic and industrial partners, the DSI is establishing its role as an international hub in data science.

Organsied by Imperial's Centre for Continuing Professional Development
www.imperial.ac.uk/cpd