DOCTORAL (PHD) STUDIES

COURSE UNIT DESCRIPTION

Course unit title	Scientific areas	Faculty	Institute, department
Multidimensional data visualization	Informatics (N 009)	Faculty of Mathematics and Informatics	Institute of Data Science and Digital Technologies

Study method	Number of credits	Study method	Number of credits
Lectures	1 (spring)	Consultations	1
Individual works	4	Seminars	1

Summary

We constantly face with multidimensional data in techniques, medicine, economics, ecology and many other areas. In fact, there are no human activities that do not accumulate and analyse such data. When developing technologies, evolving computers and software, the volume of data is growing rapidly. The needs and benefits of the right conclusions are also growing. Multidimensional data visualization is an important data analysis tool that helps to understand the structure of multidimensional data – formed clusters, highly dispersed objects (outliers), and the proximity of objects in the context of their entirety. Ph.D. students will gain knowledge of multidimensional data visualization techniques and the ability to apply data analysis tasks using popular data visualization systems.

Main topics:

- Visualization goals and its benefits in data analysis.
- Direct visualization methods: scatter plot matrix, parallel coordinates, Andrew's curves, Chernov's faces, and others.
- Projection methods: differences between linear and nonlinear projection methods.
- Principal component analysis (PCA): the nature of the principal components, principles of their computing, and the application of PCA to visualize multidimensional data.
- Multidimensional scaling (MDS): MDS principles, MDS types, similarity measurements, interpretation of results.
- Minimization algorithms for multidimensional scaling: gradient methods, branch and bound method, and others.
- Artificial neural networks for data visualization: multilayer perceptron, autoassociative neural networks.
- Self-organizing maps (neural networks): their principles, learning algorithm, additional visualization methods, interpretation of results.
- Visualization systems: Matlab tools (dimensionality reduction toolbox), Orange (https://orange.biolab.si), R package, Jupyter (https://nbviewer.jupyter.org/), DAMIS (http://damis.lt), TABLE (https://www.tableau.com), etc.
- Visual analysis applications.

Practical assignment: select a data set, visualize it using a various visualization methods, prepare a report in which the results should be described, conclusions should be drawn.

Main literature

Borg, P. Groenen. Modern Multidimensional Scaling: Theory and Applications. Springer-Verlag, New York, USA. 2005.

Dzemyda, G.; Kurasova, O.; Žilinskas, J. Multidimensional Data Visualization: Methods and Applications. Springer, 2013, ISBN 978-1-4419-0235-1. doi:10.1007/978-1-4419-0236-8.

Reddy, G. T., Reddy, M. P. K., Lakshmanna, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. Analysis of dimensionality reduction techniques on big data. IEEE Access, 8, 54776-54788, 2020.
Rice, I. Improved data visualisation through nonlinear dissimilarity modelling. Pattern Recognition, 73, 76-88, 2018.
Han, J.; Pei, J.; Kamber, M. Data mining: concepts and techniques. Elsevier. 2011.

Lecturer(s) (name, surname)	Science degree	Main publications
Gintautas Dzemyda	habil. dr.	http://www.elaba.mb.vu.lt/dmsti/?aut=Gintautas+Dzemyda
Olga Kurasova	dr.	http://www.elaba.mb.vu.lt/dmsti/?aut=Olga+Kurasova
Julius Žilinskas	(HP) dr.	http://www.elaba.mb.vu.lt/dmsti/?aut=Julius+Žilinskas