

Heterogeneous Data and Knowledge Processing

(Gabor Bella et Yannis Haralambous)

This course introduces students to the challenges and opportunities presented by data heterogeneity in the digital ecosystem. It emphasizes the importance of understanding data significance from both human and computational perspectives, combining techniques from computer science and artificial intelligence.

Key concepts covered:

- Management of diverse data types (structured, unstructured, textual, numerical, images, sensor data)
- Domain knowledge integration in data processing
- Formal knowledge management techniques
- Machine learning approaches for heterogeneous data
- Data integration pipeline development
- Formal and unambiguous representation of heterogeneous data sources

By the end of this course, students will be able to:

- [BC-04] Develop formal and unambiguous representations of complex, heterogeneous datasets
- [BC-04] Apply both machine learning and formal knowledge management techniques to process diverse data types
- [BC-07] Design and implement data integration pipelines for heterogeneous data sources

Prerequisites :

- Being familiar and efficient with Python programming
- Understanding of machine learning fundamentals, particularly classification