UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Empowering AI through Enriched Data: Combining Knowledge Graphs and Machine Learning

2526-114R-2-03

Title

Empowering AI through Enriched Data: Combining Knowledge Graphs and Machine Learning

Teacher(s)

Matteo Palmonari

Language

English

Short description

Artificial Intelligence (AI) applications largely rely on data, which often come in diverse semi-structured and unstructured formats. This course provides an overview of methodologies designed to improve the quality and utility of data that underpin AI systems through the use of Knowledge Graphs (KGs).

The course begins with an introduction to KGs as data models for knowledge representation and reasoning.

We then examine how these data models, when combined with Machine Learning (ML), can provide effective solutions for data enrichment in downstream AI applications. Two main families of enrichment paradigms will be considered: (1) semantic data integration, and (2) KG embeddings and link prediction.

(1) Semantic data integration

Semantic data integration refers to the process of constructing entity-centric knowledge repositories that integrate information from heterogeneous sources. The course will focus primarily on this paradigm, presenting a range of ML techniques for interpreting, annotating, interlinking, and augmenting tabular and textual data. These techniques include KG-based NLP methods such as semantic table interpretation and annotation, named entity reconciliation and linking, and entity clustering.

Building on these foundations, we will explore how semantic data integration can be applied in prominent downstream applications, including Question Answering systems and Retrieval-Augmented Generation (RAG) frameworks.

(2) KG embeddings and link prediction

Finally, the course will briefly introduce the second paradigm, which focuses on learning embeddings of KG representations and using them for link prediction. In this context, predicted links are understood as inferred triples derived from latent features rather than from explicit reasoning rules.

CFU / Hours

2CFU / 16 hours

Teaching period

June-July 2026

Sustainable Development Goals

INDUSTRY, INNOVATION AND INFRASTRUCTURE