KEN4256

Course Overview

Course Learning Objectives

- Able to **define** and **describe** what a Knowledge Graph is
- Able to **identify** and **describe** the **components** of a Knowledge Graph
- Able to distinguish between **different representations** of Knowledge Graphs, and **identify their strengths and weaknesses**
- Able to describe and execute **approaches to construct** Knowledge Graphs from *structured and unstructured sources*, across different domains
- Able to **construct and query Knowledge Graphs** to answer questions about their content using open standards such as RDF and SPARQL
- Use **Large Language Models** to construct knowledge graphs, and to retrieve their contents.

Learning objectives (cont.)

- Able to **execute link prediction** and associated graph mining techniques to enrich information in Knowledge Graphs
- Able to **describe the FAIR principles** and **construct KG metadata** using available standards
- Able to describe KG quality metrics and **evaluate the quality of a KG**
- Able to **develop your own KG solution for a problem of interest**

KEN4256 - 2024 Time Table

Week	Monday (C2.007)					Wednesday (C1.015)		Sunday
WK1 Feb 5 - Feb 9	13:30-14:30 (LE1) Intro to KG	14:30-15:30 Course Overview, LLM Ethics Projects	15:30-16:00	16:00-17:00 (LE2) KG Construction from Structured Data	17:00-18:00 (LA1) Exploring Wikidata	16:00-17:00 (LA2) Construct RDF graph with python	17:00-18:00 (LA2) Construct RDF graph with python	
Carnival Feb 12 - Feb 16								
	BLK1 11:00-12:00	**BLK2** 12:00-13:00	**BREAK** 13:00-13:30	**BLK3** 13:30-14:30	**BLK4** 14:30-15:30	**BLK1** 16:00-17:00	**BLK2** 17:00-18:00	Due
WK2 Feb 19 - Feb 23	(LE3) KG Construction from Unstructured Data	(LA3) KG from Unstructured Data		(LA4) Semantic interoperability with shared vocabularies	Project proposal writing. Introduce Assignment 1.			Project Proposal
WK3 Feb 26 - Mar 1	(LE4) KG Retrieval (SPARQL)	(LA5) graph queries with SPARQL		(LE5) KG Quality	(LA6) Quality (sparql queries)	Work on Assignment 1 & Project	Introduce Assignment 2	Part 1: semantic interoperability
WK4 Mar 4 - Mar 8	(LE6) Publishing KGs	(LA7) Graph Metadata		Work on Assignment 2 & Project	Introduce Assignment 3			Part 2: Data Quality Assessment
WK5 Mar 11 - Mar 15	(LE7) Graph Embeddings	(LE8) Link Prediction and Explainable AI		(LA8) GNN - link prediction & explanation	(LA8) GNN - link prediction & explanation	(LE9) KG Analytics and Visualization	(LA9) Graph analytics & Visualization	
WK6 Mar 18 - Mar 22	(LA10) RAG over KG	(LA10) RAG over KG		Take up Assignments 1 & 2.	Work on Project Report and Course Wrap Up.			Part 3: RAG over KG
WK7 Mar 25 - Mar 29								Project Report

Course materials

- Slides for the labs and lectures will be released on **Canvas** just before the respective session in **PDF** format. You can download and follow along during lectures and labs.
- Assessments will also be released in **PDF** format on **Canvas** according to the dates indicated in the previous slide for evaluation.
- Your solutions for the assignments and the project will also be uploaded on **Canvas** on the deadlines indicated on the previous slide. More instructions will be provided in the description documents for these assessments when they are released.
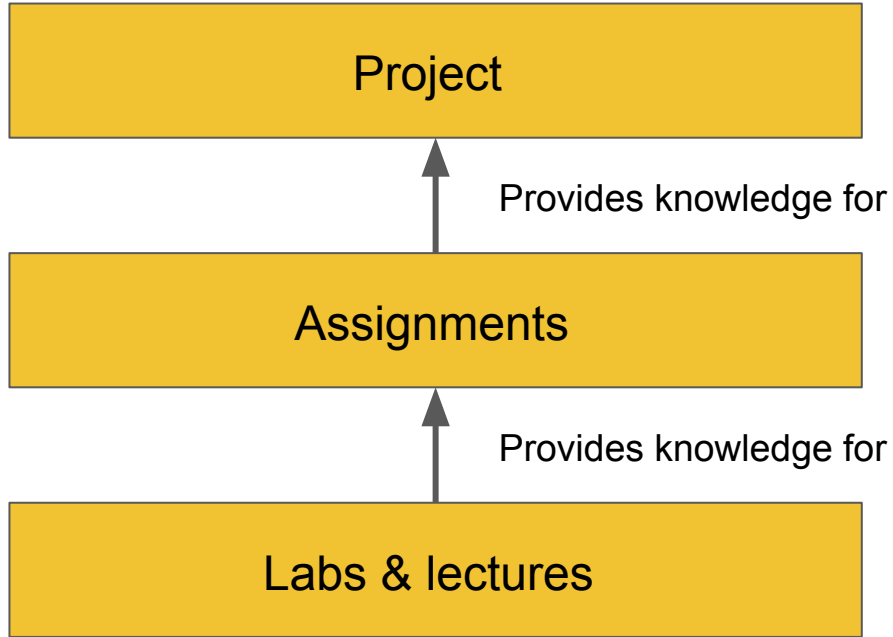
Evaluation

- **3 Group assignments** to <u>build knowledge and core skills</u> (45% of final grade; 15% each)
  - Assignment 1
    - Semantic interoperability over structured and unstructured data
    - Released **February 19**, due **March 3 (23:59 CET)**
  - Assignment 2
    - Data quality assessment with wrong data
    - Released **February 26**, due **March 10 (23:59 CET)**
  - Assignment 3
    - RAG over knowledge graph
    - Released **March 11**, due **March 24 (23:59 CET)**

- **1 Individual project** to demonstrate <u>your creativity and technical virtuosity</u> (55% of final grade; 10% project proposal, 45% written project report)
  - Project description released **February 12**
  - 2-page project plan: **February 25 (23:59 CET)**; feedback within 1 week.
  - Final project report: **March 31 (23:59 CET)**

no final exam!

The AI/LLM space is moving fast...

- We'll be using LLMs for KG construction, query answering (RAG), and potentially other uses.
- Recommend using Open Source LLMs such as Mistral, LLAMA, etc. Check out the open source LLM leaderboard from hugging face. new LLMs make the top every week.
  https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard
- Many tools for working with LLMs (CLI, RAG, others) include Ollama, Langchain, LLMware, LLamaIndex
- Learn what the community is doing and get inspired for your project.
- Ideally, you have a GPU powered laptop. Otherwise you may consider a Google colab (pay as you go & subscription plans)

Recommended reading

https://arxiv.org/abs/2003.02320

Knowledge Graphs

AIDAN HOGAN, IMFD, DCC, Universidad de Chile, Chile
EVA BLOMQVIST, Linköping University, Sweden
MICHAEL COCHEZ, Vrije Universiteit and Discovery Lab, Elsevier, The Netherlands
CLAUDIA D'AMATO, University of Bari, Italy
GERARD DE MELO, Rutgers University, USA
CLAUDIO GUTIERREZ, IMFD, DCC, Universidad de Chile, Chile
JOSÉ EMILIO LABRA GAYO, Universidad de Oviedo, Spain
SABRINA KIRRANE, SEBASTIAN NEUMAIER, and AXEL POLLERES, WU Vienna, Austria
ROBERTO NAVIGLI, Sapienza University of Rome, Italy
AXEL-CYRILLE NGONGA NGOMO, DICE, Universität Paderborn, Germany
SABBIR M. RASHID, Tetherless World Constellation, Rensselaer Polytechnic Institute, USA
ANISA RULA, University of Milano–Bicocca, Italy and University of Bonn, Germany
LUKAS SCHMELZEISEN, Universität Stuttgart, Germany
JUAN SEQUEDA, data.world, USA
STEFFEN STAAB, Universität Stuttgart, Germany and University of Southampton, UK
ANTOINE ZIMMERMANN, École des mines de Saint-Étienne, France

In this paper we provide a comprehensive introduction to knowledge graphs, which have recently garnered significant attention from both industry and academia in scenarios that require exploiting diverse, dynamic, large-scale collections of data. After some opening remarks, we motivate and contrast various graph-based data models and query languages that are used for knowledge graphs. We discuss the roles of schema, identity, and context in knowledge graphs. We explain how knowledge can be represented and extracted using a combination of deductive and inductive techniques. We summarise methods for the creation, enrichment, quality assessment, refinement, and publication of knowledge graphs. We provide an overview of prominent open knowledge graphs and enterprise knowledge graphs, their applications, and how they use the aforementioned techniques. We conclude with high-level future research directions for knowledge graphs.

CCS Concepts: • **Information systems → Graph-based database models; Information integration;**

Additional Key Words and Phrases: knowledge graph

1 INTRODUCTION

Though the phrase "knowledge graph" has been used in the literature since at least 1972 [465], the

Recommended reading

- **A Semantic Web Primer**. 3rd Edition. Grigoris Antoniou, Paul Groth, Frank van Harmelen and Rinke Hoekstra. 2012. MIT Press, ISBN: 9780262018289.
- **Semantic Web for the Working Ontologist.** 3rd Edition. James Hendler, Fabien Gandon, Dean Allemang. 2020. Morgan Kaufmann. ISBN-13: 978-1450376174; ISBN-10: 1450376177.
- **Practical RDF**. Shelley Powers. 2003. O'Reilly Media, Inc. ISBN: 9780596002633
- **Learning SPARQL**. Bob DuCharme. 2011. O'Reilly media, Inc. ISBN: 9781449306595
- **Programming the Semantic Web**. Toby Segaran, Colin Evans, Jamie Taylor. 2009. O'Reilly Media, Inc. ISBN: 9780596153816